

**OPERATIONAL RECORDS:
THEIR NATURE AND POTENTIAL USE
IN A RESEARCH CONTEXT**

Pascoe Pleasence

1999

OPERATIONAL RECORDS: THEIR NATURE AND POTENTIAL USE IN A RESEARCH CONTEXT

Introduction

This chapter is intended as a guide to the nature, potential use of and potential problems associated with operational records in a socio-legal research context. As the issues raised have a general bearing, the examples are largely set out in hypothetical form. Although some operational recording systems are more useful in particular research contexts and some organisations allow greater access to their records than others, these are principally matters of degree.

In the context of this chapter *operational records* are taken to cover the entirety of data collected by organisations (or individuals acting in a business capacity) in the course of undertaking their *primary functions* - and then stored.¹

The reason for singling out primary functions is simply to exclude ancillary research functions and their associated data systems, which will clearly be set up in accordance to different considerations.² This chapter will not, therefore, address issues relating to the sharing of research data or co-ordination of strategic research projects. Whilst interesting and extremely important areas, they are for another occasion. It will, though, address the issue of data compatibility between organisations operating in overlapping fields.

The chapter starts with a brief overview of the range of operational records. It then explores, in four main sections (Design/Function, Responsibility/Reliability, Sensitivity/Confidentiality, Quantity/Detail), the many tensions that exist in creating, operating, maintaining and using the data from operational recording systems. It concludes by offering some thoughts on how researchers might best regard operational records and how they should go about using them.

The Range of Operational Records

There are vast quantities of operational records relating just to the justice system. They lurk on desks, in drawers, in cupboards, in warehouses, on personal computers on mainframe computers and even in people's heads.

¹ The fact that data is stored is no guarantee of it being stored for any great length of time. Most organisations periodically destroy paper files, often after the end of a regulated storage period. Some organisations also decommission electronically stored data, especially in cases where the systems the data is stored on are superseded by newer, but incompatible, systems.

² As will be seen, though, many of the issues raised below touch upon all recording systems.

Those operational records which researchers are generally most familiar with, and which are often seen incorporated into research projects to date, include:

1. lawyers' case files – which include records of work undertaken, case notes, correspondence items, statements, reports, application forms for funding, items of evidence, and, generally, case summaries and progress forms;
2. lawyers' management systems – which can include diary, work activity and billing information;
3. court documents – ranging from applications received by courts to evidence lodged with them;
4. court records – which include details of who appears before the courts, the reasons for appearances and administrative information relating to their functioning;
5. taxation records;
6. funding authority records (e.g. those held by the Legal Aid Board, trade unions and insurance companies) – which contain client, supplier, case and costs information;
7. police and prosecution records – which may be held by organisations as diverse as the police, the crown prosecution service and local authorities;
8. central governmental records (e.g. those held by the Home Office, Lord Chancellor's Department and Scottish Office) – which contain details of the administration of the justice system, as well as highly specific data such as the Home Office's offenders' index.
9. professional regulatory records (e.g. those held by the Law Society and Bar Council) – which include, for example, membership and accreditation information.
10. Non-lawyer legal service supplier records (e.g. those held by mediators and non-lawyer advice agencies).

Given this range of potential research data it might be possible to forgive someone for thinking we live in something of a socio-legal research paradise! However, these are sentiments that would necessarily be shared by those who have experience of using operational records within their research. Operational records do not always deliver all that they promise, and what they can promise is itself often limited in scope and form.

System Design and System Function (or System Design and System Use)³

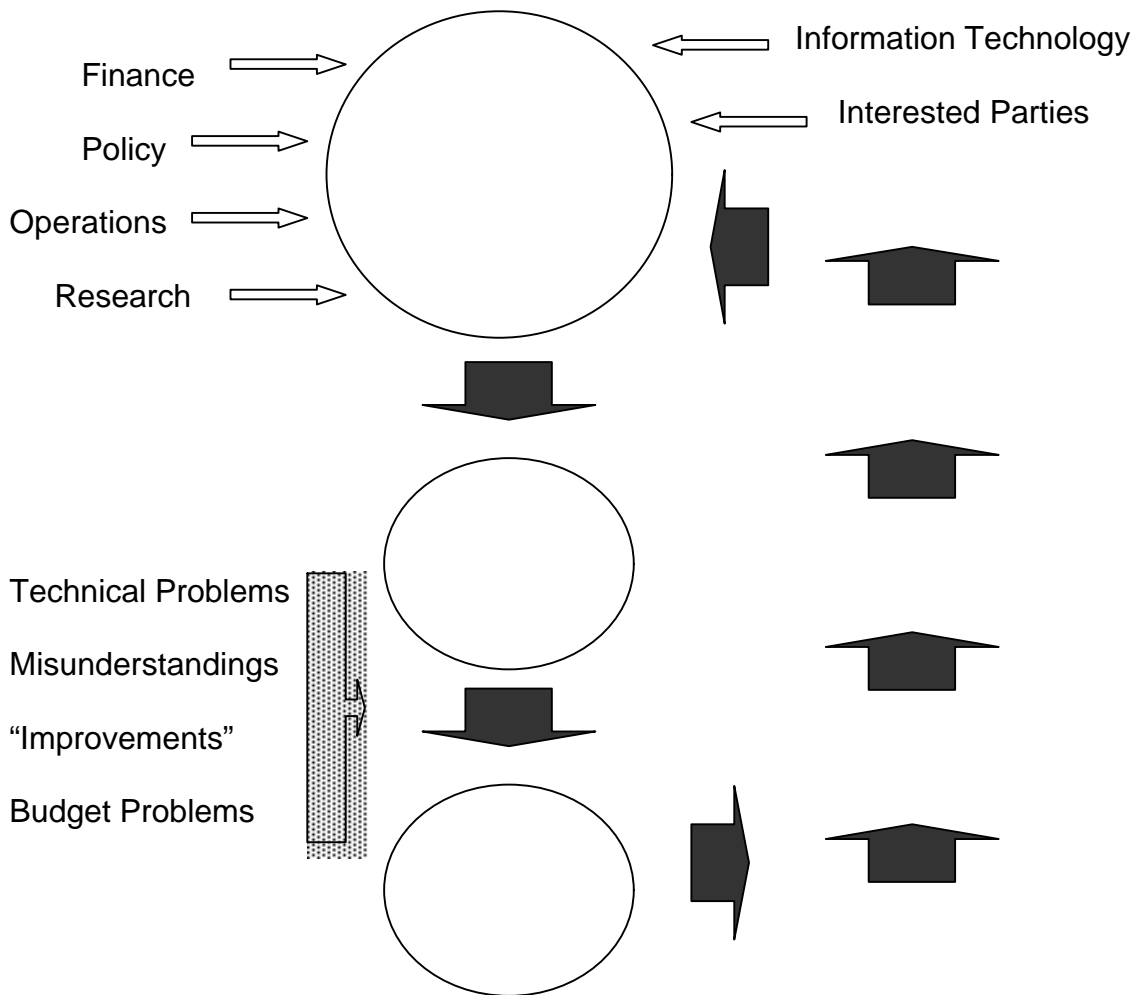
It may come as a surprise to some, but operational recording systems are not put together on the basis of arbitrariness. The grand computer based operational data systems employed by large organisations, for example, are generally

³ For the purposes of illustration this section is focused on the operational recording systems of large organisations. The discussion does have application to the workings of small organisations and individuals, though, as will be seen, a good number of the points made have particular bearing upon large and functionally fragmented organisations.

designed at immense cost, with a great amount of internal (and often external) consultation (especially at the early stages), and updating to concerned and interested parties throughout.

A typical design process looks something like this.

Figure 1
Basic Design Process for Operational Computer Based
Operational Data System



Although there is often a great deal of effort spent drawing up the system specification for a new computer system and vast sums spent on building the system, the system rarely meets all of the needs of those who will ultimately use it. The first problem is that a lot of people may need to have their say, and a good number of them have very different interests.

Those holding the purse strings may not want to hand over their coin, so from their perspective the system should be as streamlined as possible. Those at the face of operations do not want to overburden their staff and are constantly mindful of the targets that they must meet in actually delivering whatever it is the organisation delivers, so they may share the less is more philosophy. They may also be concerned that information be presented to their staff and inputted onto the system in a manner which is understandable to their staff (who may be more or less well trained in the business of the organisation).

Those with policy interests want to ensure that their specific goals are measurable and that data is made available to inform the future development of their ideas. They may want changes to the way data is presented and stored. They may want a great deal of data to be stored. For them more is more and new may be better than old. Those with research interests may be even more data hungry, wanting to guarantee their tools exist for long term study and analysis. They also may want changes to the way data is presented and stored.

Those who are concerned with providing specific information in organisational reports may want to see compatibility with past data. As may external bodies and individuals for whom the reports are produced.

Those building the system may naturally want to build it in a way they are comfortable with. They inevitably also bring with them limitations of technology and limitations of competence.

So, from the start operational recording systems may be the subject of great compromise, even from the purely operational perspective.

After initially determining what the system must do, the system builders start to build. They will generally keep interested parties informed, but this may not always be in a user-friendly form, and there is always a danger that everybody is under the impression they share an understanding of what the system should look like and so take less effort to study the building process.

Technical problems will inevitably be encountered at the initial build stage. "Improvements" may be made by the builders on the basis of their particular perspective and knowledge. One person's improvement may be another's nightmare! Misunderstandings can be amplified, and budgetary pressures may force corners to be cut. The initial design may mean that the corners that can be cut are not those that would ideally be cut.

Finally systems go live. Hopefully they work, but even when they do the practicalities may mean that they turn out to be too cumbersome. System inadequacies may mean data entry takes too long. The system may consequently need to be re-worked. At this stage, though, there may be panic in the air. Those closest to the system building and system use (i.e. those charged

with data entry) may make snap decisions to cut what is easily cut. Again, if a system is not particularly flexible, the easy cuts may be the most unfortunate. There is, of course, always some scope for the design process to be cyclical, but pragmatism and practical reality generally see those closest to a *problem* simply *solve it* to their satisfaction.

Always, in any event, and at every stage, inequalities in the bargaining power of the various interested parties play a massive role in the design and operation of operational recording systems. Not all the arrows in figure 1 carry equal weight. Also, in an operational context, it is often regarded as more important that something works than that it works well. Much of the time this is simple pragmatism, but it can also be the manifestation of a need to be protected from charges of failure as defined by targets. However, as will become evident below, meeting targets, working and working well are all very different things.

One last point that must be made in the context of operational recording system design is that sometimes, during the often lengthy design process, the whole world changes and the operational needs of the system change unrecognisably. It then has to be hurriedly adapted, perhaps to tasks for which it is ill prepared or ill designed. This may mean the form, range and quantity of data is compromised to an even greater extent.

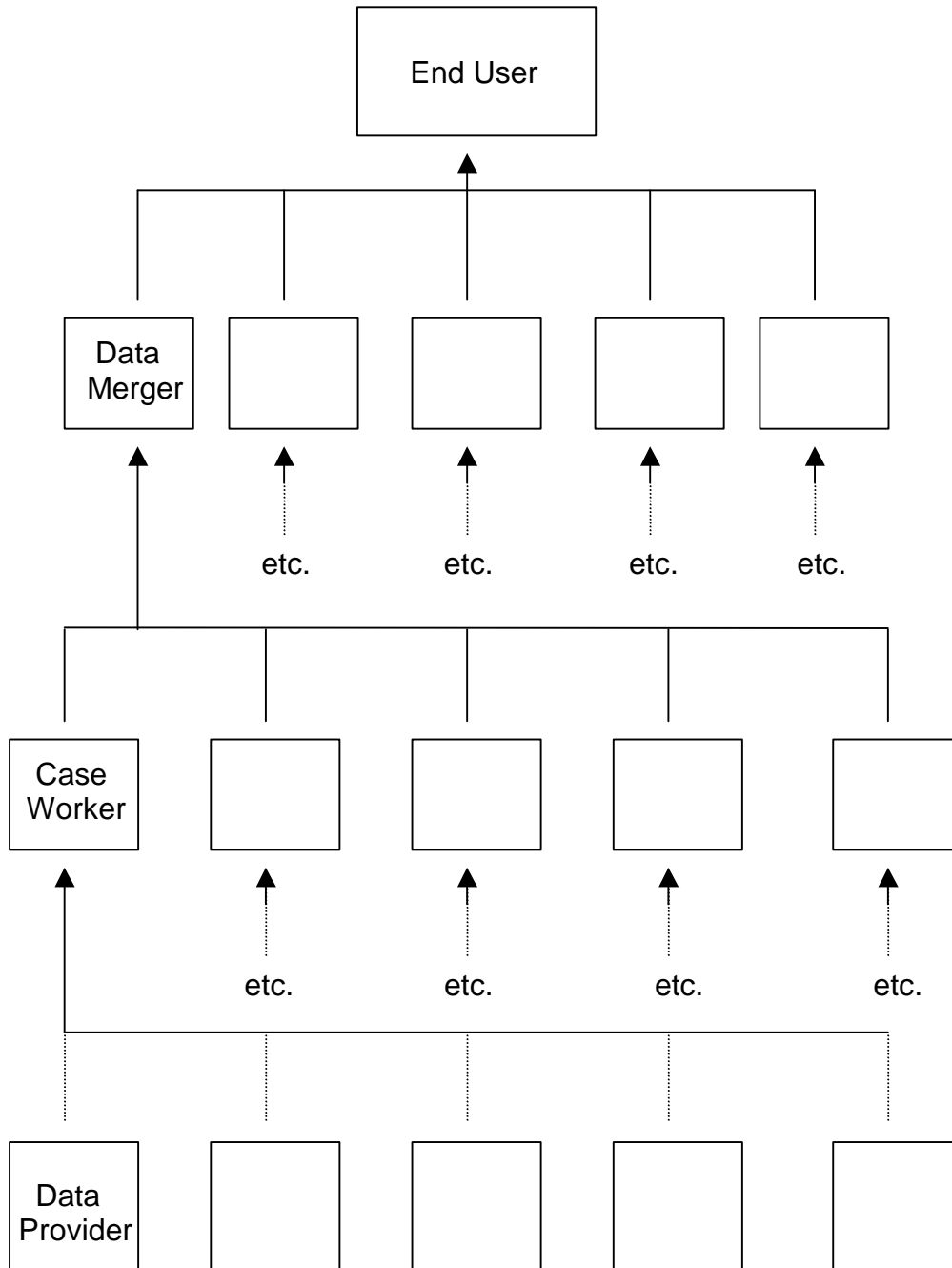
Responsibility and Reliability

Continuing for a while with this picture of the large operational data system, let us turn away from the system itself and look at the organisation.

Those who actually undertake the day to day data entry in organisations, both large and small, may be very far removed (sometimes physically, managerially and socially) from the data system design. They may also be very far removed from the ultimate data use.

Figure 2 provides a simple illustration of a hypothetical information. At the top are the data users (e.g. policy makers), at the bottom are the data providers (e.g. individual lawyers). In between are various data transcribers and transformers. There are those who prepare the information collected by the lawyers for transmission to *the organisation*. There are the caseworkers who receive the information and input it into the organisational systems. There are the processes by which information from the various caseworkers and operational systems are combined for the benefit of the ultimate users. The ultimate users ultimately report upon the data to others.

Figure 2
Management Structures and Information Paths



The first point to notice in looking at figure 2 is that there are a lot of potential error points. Those providing the information may provide it in an erroneous form. Those receiving it may erroneously record it. Errors may creep in where data is moved from one data system to another (maybe from remote locations to a central system or from independent systems dealing with different aspects of

operations). Errors may also be introduced at the data analysis stage. In addition to this the data may be misunderstood and incorrect analyses consequently derived from it.

The forms of error and the reasons for errors creeping into an information chain may be numerous. There may be a lack of training as to correct forms of, say, data entry. There may be carelessness or even incompetence. There may be bias. People at different points in the chain have different interests as regards the look of the data. There may be short-cutting. There may be translation errors.

Bias is a particular problem where the data supplier is dependant in any way upon the form of the data. So, for example, if the form of the data determines the level of payment for a service there may be a tendency, on the part of the data supplier, to err in a manner favourable to them.

Short-cutting can be a major problem in large organisations.

As already noted, different parts and different levels of an organisation may be working to very different agendas. Different organisations may be working to agendas more different still.

An interesting example is provided by the data entry caseworker who is working primarily to strict volume targets under the management of someone working to similar targets, but multiplied up.

Such caseworkers, have probably not been fully advised of the meaning, use and importance of the data they process. They are even less likely to have been briefed on the relative importance of different data elements they deal with. They also probably have little contact with those who ultimately use the data or design the systems they work within.

It must be expected that if the quality of data entry comes into conflict with that of a caseworkers principal target, then there is a real danger that the data will be compromised to achieve the target. In a large organisation which has a continuous and extensive supply of new data that requires entering onto the organisations systems and slippage in meeting targets may rapidly lead to vast back-logs, great managerial pressure and great pressure from outside the organisation (especially where the data processing is critical to the functioning of external organisations).

Data compromising may take a number of forms. First, it may simply involve missing out the entry of specific data elements. This may be easy to spot, but it need not necessarily be so – especially if the ‘truth’ is manipulated to make it appear that all data has been correctly entered. For example, if there are a number of related passenger claimants following a car crash and data is entered as if there was only one, to avoid filling in multiple claimant details (etc.), then

there may be no way of determining that data is missing in respect of that entry without recourse to source data.⁴ Error correction systems and the use of mandatory fields may perversely conspire to worsen the situation. If the only way to take a short cut through a system is to enter a corrupt piece of information, because the information must be entered and if it is entered correctly an additional set of data will need to be entered, then corrupt data at that critical point in the system is made more likely.

Second, it may involve putting dummy data into the system. This may be to avoid following up gaps in source data. It may be because of data ambiguity. It may involve an alteration to the data to realise the time gain that a standard entry may promote. It may involve a sensible *guesstimate*. It may involve arbitrariness. Some dummy data are easy to spot (e.g. dates of birth of 01.01.01), but some may be almost impossible to spot at the individual case level.

Caseworkers may not be the originators of *sharp practice*. Managers can face the same pressures and instruct caseworkers to adopt short-cutting practices. Sometimes these practices will be brought to the attention to those higher up the management ladder, sometimes not. Sometimes there will be top down instructions to foreshorten the data entry process. Generally the higher up the ladder a decision is made the more likely it will be brought to the attention of those who use the data, but this need not always be the case.

If caseworkers or managers are not aware of the use to which data will be put or its importance, then there is even less of a barrier to its corruption.

If some of the hands through which data travels are in different organisations the problems can become extreme. This is more and more so as the leverage of one organisation over another becomes less and less. If the functions and responsibilities of an organisation are seriously fragmented the problems can become extreme. If people processing data are at a great distance (in responsibility terms) from those who use it, the problems can become extreme.

Ultimately, if data sets become seriously corrupted through the frequent inclusion of inaccurate information it is important to ask the question whether the data is worth collecting at all. Nonsense data is very expensive and of little use to anybody.

It may be that in today's climate it may be sensible for many organisations to collect only minimal amounts of information on a universal basis. Many data hungry functions perhaps ask too much of universal systems. Guaranteed systems with a flexible ability to collect detailed information on a sample basis,

⁴ It may be apparent at the macro level that there is systematic avoidance of certain forms of data entry, but it will be extremely difficult to separate the real from the apparent single defendant cases.

which can link to and expand upon universal core information would seem to be more sensible than monolithic information systems.

An end note here would be that *some* computer held information in large organisations is very reliable. Organisations, for example, are generally well aware of how much money they pay out and to who. The further away from core function data is, the more prone it is to error – unless it is collected in a dedicated system with care taken over quality assurance.

Responsibility and Compatibility

A problem frequently encountered by researchers using electronically stored operational records is that data held by different organisations cannot always be linked together.

As was shown above, a great many organisations, of many sizes, form a part or have an involvement in the justice system. Some of them are on what might be described as *opposite sides* and have understandable reasons for not wanting to devote a great deal of time to harmonising their data recording methods. But even those on the same side often have less than harmonious data systems.

An interesting example here is the cost of defending criminal cases. Most defence costs are met through legal aid expenditure. However, there is currently no single body that controls even the payment of legal aid fees. Until cases leave the Magistrates' Courts lawyers look to the Legal Aid Board. In the Crown Court they look to the courts and the Lord Chancellor's Department, perhaps via the National Taxation Team. No institutional body has an operational records system that spans the whole process.

If institutional records are to be used to make a study of defence costs in criminal cases, therefore, operational recording systems from more than one organisation will need to be examined. In the case of legal aid defence costs it is not currently possible to link costs incurred in the Magistrates' Courts with those incurred in the Crown Courts. Although the Magistrates' Courts and Crown Court cost data is not fundamentally incompatible⁵, there is no unique case identifier which will allow the linking of records held on the relevant different systems.

The only ready explanation for this would seem to be that different organisations tend to confine their interest and efforts to those matters which fall within their responsibility. Another key problem is simply that different organisations rarely change their data systems at the same time. Thus there is rarely any unified and actionable interest in co-ordinating data systems. Further, there is unlikely to be an underlying unwillingness to incorporate *linking* data into a system where it is

⁵ There are, though, differences in the classification systems (e.g. offence type codes) used and the items of data stored.

perceived as being of no direct use to the organisation that will be paying for its collection. Every keystroke at a computer has a cost attached and avoidable cost is the first to be chopped.

If any body or institution wishes to gain an overview in an area of the justice system where function and responsibility is fragmented, then they will have to convince those who could collect the information they require that it is worth their while. They will also, most likely, have to pay for its collection.

Quantity and Detail

The more detailed we require our data to be, the more difficult it is to obtain. Large operational computer data sets, for the reasons indicated above, can contain only limited data in relation to individual matters. They may contain mountains of data in total, but they generally lack depth.

To obtain depth it is necessary to further and further back down the information trail, as illustrated in figure 2, and ultimately to go to the working files kept by the various players in the justice system (and even to players themselves).

The most extensive files to be found in the justice system are those kept by lawyers, whether in private practice, prosecuting authorities or other organisations. They are also pretty reliable.

However, working through files is an extremely time consuming (and soul destroying) exercise. To complicate matters, different files are organised in different ways. Some also employ elaborate coding schemes (and almost all will, in respect of fee earners etc.). Unfortunately, some also are barely legible. However, on the whole the lawyer's file represents a rich mine of information.

Working through ten files a day, for data collection purposes, is pretty good going. Thus, to obtain a sample of 1000 cases, half a year of sitting with files may be called for. When account is taken of making arrangements to view files, travelling to where they are kept (if necessary), and developing the collection process, it is clear that many more months are usually required.

On top of this, if the files are to be randomly sampled, great care must be taken over selecting both lawyers and files. Truly random sampling is perhaps unattainable. Many lawyers will not grant access. Many files will be lost, missing or held back. Misunderstandings will arise over what constitute "files", "cases", etc.. Different lawyers will adopt different practices as regards when files are opened and how many files are opened in respect of each client or case or whatever it is that is key to a filing system. Researchers must, if they want to treat information drawn from many different lawyers' files together, develop units of measurement which are compatible across all of the files surveyed.

Furthermore, files are not devoid of inaccuracies. Carelessness can strike anywhere, as can bias, as can honest misrepresentation. Lawyers cannot be expected to make a note of all the work they do in respect of a case, nor can they be expected to be 100% accurate in respect of those notes they do make.

In contrast to the great difficulty that can be experienced in collecting data from files, collecting data from the large organisational computer systems can be a breeze. For example, with one swift 2 minute telephone call it would be possible to obtain data on many hundreds of thousands of legally aided cases from the Legal Aid Board. The data could relate to who brought them, how much they cost and what the basic outcome was.

Nothing could be found out about the strategy adopted by the parties to the case, of course, nor the detail of the work undertaken, nor the reason for delays, nor the reason for the recorded outcome.

Determining exactly what information is truly required and exactly what information is available and how it can be obtained is the most important part of any research project. A small compromise may save many months. A small error may cost the same.

Sensitivity and Confidentiality

As those who have sought to obtain research data from commercial concerns will testify, they are often extremely reluctant to part with information which goes to the core of their business. Insurance companies are not always happy to part with the information on which they base their risk analysis, that contains their client details, that offers them routes to commercial advantages over their competitors, and which may contain confidential disclosures. The more sensitive information is, the greater the hurdles that may need to be overcome to negotiate access to it.

So far there has been much discussion about where to find data, how it got there, and how good it is. However, there has been little mention of the crucial fact that researchers cannot simply obtain information simply because they need it.

Much of the information in which we are interested as researchers also relates to individuals' personal characteristics. Information which they will often have provided in confidence. Information which is confidential, may also be commercially sensitive, and may, in some circumstances, be wholly unlawful to divulge.