CONFERENCE OF EUROPEAN STATISTICIANS
Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality Session 5
29-31 October 2019, the Hague, the Netherlands 30 September 2019

Understanding personalities in data access decision-making

Richard Welpton (The Health Foundation, United Kingdom)
richard.welpton@health.org.uk

Abstract and Paper

In the UK, a number of Safe Settings now exist which enable secure access to detailed, confidential data in the health and social science sectors. Accessing these data have allowed researchers to generate important insights about our society. The Safe Settings through which they are accessed have proven robust for protecting their confidential nature.

Accreditation for Safe Settings has existed for some time. Successive waves of legislation introduced in the previous decade have increasingly provided gateways to enable access to confidential microdata for research purposes. Often, Safe Settings provide comprehensive training to researchers before can access the data. There have been few, if any, serious breaches of data access from a Safe Setting that have resulted in the confidentiality of data being broken.

These are positive developments. However, many data suppliers remain reticent about giving the go-ahead to allow the data for which they are responsible to be accessed, even in a Safe Setting. Often the decision is made by one person (an Information Asset Owner), who may be receptive to hearing about the benefits of data access, and they may be assured by the safeguards offered by a Safe Setting. However, they attache more weight on the risk of an adverse event occurring. They naturally default to a prolonged no-access decision.

This paper explores the choices and behaviours of Information Asset Owners to consider how better informed and timely decisions can be made about access to detailed and confidential microdata. Ultimately, understanding the mindset of such an individual is the key to unlocking access to such data in the future.

Understanding personalities in data access decision-making

Richard Welpton*
*The Health Foundation, richard.welpton@health.org.uk

Abstract

In the UK, a number of Safe Settings now exist which enable secure access to detailed, confidential data in the health and social science sectors. Accessing these data have allowed researchers to generate important insights about our society. The Safe Settings through which they are accessed have proven robust for protecting their confidential nature.

Accreditation for Safe Settings has existed for some time. Successive waves of legislation introduced in the previous decade have increasingly provided gateways to enable access to confidential microdata for research purposes. Often, Safe Settings provide comprehensive training to researchers before can access the data. There have been few, if any, serious breaches of data access from a Safe Setting that have resulted in the confidentiality of data being broken.

These are positive developments. However, many data suppliers remain reticent about giving the go-ahead to allow the data for which they are responsible to be accessed, even in a Safe Setting. Often the decision is made by one person (an Information Asset Owner), who may be receptive to hearing about the benefits of data access, and they may be assured by the safeguards offered by a Safe Setting. However, they attache more weight on the risk of an adverse event occurring. They naturally default to a prolonged no-access decision.

This paper explores the choices and behaviours of Information Asset Owners to consider how better informed and timely decisions can be made about access to detailed and confidential microdata. Ultimately, understanding the mindset of such an individual is the key to unlocking access to such data in the future.

Introduction

In recent years, the environment for accessing confidential, sensitive data about individuals for the purposes of analysis and research has improved significantly. The environment has been aided by changes in legislation, and to developments in technology.

The introduction of the Statistics and Registration Services Act in 2007 brought about the legal concept of the Approved Researcher, putting in place a legal framework for accessing sensitive data for research purposes. However, this legislation applied to underlying data that would be used to create official statistics. More recently, the Digital Economy Act (2018) provided a more comprehensive regulatory framework to enable researchers to access data from any government agency, including administrative data which might not necessarily be used to create official statistics (noting that health data are largely excluded). The DEA also enables data to be more easily shared amongst government agencies; enabling the linking of individual data from various administrative sources. Prior to the Digital Economy Act, legislation such as the Social Security Act 1991 did not reflect and therefore facilitate, use of modern technologies and statistical techniques that could use administrative data, and provide insights that would benefit society (such as policy evaluations).

These legislative developments have occurred while the number of Safe Settings for accessing sensitive data have proliferated. Services such as the Office for National Statistics' VML (established in 2004-2006), the UK Data Service Secure Lab(providing secure remote data access) routinely provide access whereby confidential data can be safely analysed. In addition, a number of universities and charities have established their own Safe Settings, including Cancer Research UK and The Health Foundation. The Turing Institute, and now University of Manchester have developed Safe Settings too.

What is a Safe Setting? This refers to the set of security controls in place to protect the data. They are technical in nature (computing and network solutions); and human (policies and procedures), accrediting and training users. Generally, these Safe Settings enable sensitive data to be accessed and

analysed in a single secure location; the data never leave the Safe Setting, but Safe Outputs (typically statistical results which are vetted to ensure they do not contain any information that might re-identify or contain confidential information) may be released.

With the combination of new legislation and Safe Settings, one might believe it is possible that more access to sensitive data can be achieved. There have been advances; the SAIL project for instance in Wales, now routinely provides access to a range of linked confidential data. The new Administrative Data Research Partnership will build upon the work of the Administrative Data Research Network to encourage and facilitate data sharing, linking and access to sensitive data. Government agencies now speak regularly about the benefits of sharing and linking data, and enabling access.

Why then, do researchers continue to experience problems in accessing such data? Often, a researcher will think of a policy evaluation problem, or a hypothesis, which can be solved undertaking some empirical analysis. The requirement for data is the foundation of such an analytical project. Identifying the right data source is crucial; and while accessing such data can be, on the face of it, straightforward where that data source has been used before (i.e. where there is precedent), the same is often not the case where novel data sources are requested.

This is a problem which has not been resolved by advances in technology and facilitation by new data legislation. In the UK, an individual known as an Information Asset Owner (IAO) ultimately makes the decision about whether data can be accessed.

The rest of this short paper considers the role of the IAO, and how their perceptions of risk shape their decision for enabling access to data, while fulfilling their job to protect the confidentiality of the data. In particular, this paper outlines some thoughts about how data access decisions made by IAOs can be made to ensure that the risk of providing access to data is minimised, but the benefits of such data can be realised.

Who is an information asset owner?

In general, data sources maintained within government agencies are 'managed' on a day-to-day basis by an Information Asset Owner (IAO). The role is typically assigned to a mid-ranking civil servant. They assume responsibility for the proper collection, storage, use, and (in some cases) disposal of the data. This includes deciding whether applications by researchers to use the data for analysis can be permitted. In principle, their decision should be guided by considerations as to whether:

- use of the data is legal
- any risks to the confidentiality of the data
- Benefits to society for enabling access to the data

Doubtless, other considerations are taken into account as well.

We might assume that all IAOs have some background relevant to their role. Experience in statistical analysis for example, information governance or data management might provide the necessary skills in order to execute their charge. While some certainly will have a relevant background, it is entirely possible that a broad experience covering both statistical analysis and data management is lacking for some. The result is that they may not have the confidence, or feel empowered, to take decisions especially when there is some element of risk.

In these situations, saying no to a request to access data for which there is no precedent might become the default option. Why might this be the case?

Consider an IAO without a suitably broad experience. They know that at they have the power to grant or deny access to a data source. The information they have is:

1 that the dataset is confidential
2 that probably some research can be undertaken, but they don't really understand what the research involves, and it could take a while for the results to be produced, and even then, they may not understand what the results mean
3 if there is any breach of confidentiality, then they as the IAO will have to take responsibility

In this situation, the IAO understands confidentiality and risk; they don't understand so well the usefulness of the data and the research that could be undertaken with the data. Therefore, when asked to make the data available, they can make an initial assessment: they are legally obliged to protect the confidentiality of the data; ascertaining whether the proposed use of the data is legal is uncertain if they cannot ascertain the benefits; and it's easy for them to conclude that the risks of providing access to the data outweighs any benefits.

Some government agencies have established processes for managing data access requests. An approval process is in place, whereby application forms are submitted by researchers. These are assessed by data access officers, ensuring that the researcher has explained why they require the data, assessing whether there is a legal basis, how the data will be used etc. After the screening is complete, the application will be handed to the IAO who will make a decision. The IAO should by now have complete information to be able to make an informed decision, however, this author has spoken to an IAO on a phone meeting to discuss an application, where the IAO was getting on to a train and they hadn't received the application. On the basis of information received, they could only deny use of the data requested.

Often a decision is made to pass the request to colleagues, including senior management, throughout the organisation. This will delay data access, and the delay could be considerable. But now, the decision of the IAO is supported by the organisation; and they may be awarded for taking a cautious approach.

This leads to another observation: are IAOs rewarded for being cautious or for promoting access to data? This author suspects the former, because the reputational consequences at least, as well as the legal consequences, of a breach of confidentiality are likely to weigh heavily in the minds of the senior management.

In short: the IAO will take the risk for enabling access to data; but if they do not perceive any benefits from doing so, why would they take the risk? This implies that the risk/reward balance is not attuned in a way that access to data can be realised.

This leads to the 'default-closed' position described by Ritchie (2016), and access to novel data sources is difficult, if not impossible. Privacy may be protected but society loses out from research not being understand.

The next section considers remedies for this.

Enabling Information Asset Owners

Despite the changing legal environment, and the available technology, and the now tested procedures and infrastructure (such as research accreditation), it remains important important to understand the position of an IAO position: taking lots of risk but little receiving little benefit.

The nature of research is that the benefits from analysing the data may not be realised for some time. However, it is easy to consider confidentiality risks; these can be thought of, possibly exaggerated, recorded, and actioned by not enabling access. In particular, if use of the data is perceived so risky that the benefits do not align with a legal basis, then access is most certainly likely to be denied.

Because of the natural gap in the production of statistical results and their impact on society, it is easy to think about the risks to providing access to sensitive data, but not consider the benefits. This is compounded by the fact that researchers do not always relay their findings back to the IAO, let alone involve them in the dissemination of results. In addition, staff undertaking the role of an IAO may move jobs, certainly before the results of an analysis are reported. And, there while it is true that risks can be realised, a not insubstantial literature exists which discusses the role of a data intruder. While plausible, and useful as a framework to discuss risks, without a balanced approach around the virtues of data access, IAOs will see giving data access as a risky activity, and revert to the 'default-closed' position.

Fortunately, frameworks do exist which that enable IAOs to take a balanced approach to dataset. The Five Safes (Desai, Ritchie and Welpton, 2016) takes a step-by-step approach for considering access to data. In particular, the framework encourages one to consider:
- is use of the data 'safe'?
- can the user be trusted with the data?
- Is the setting secure?
- Can we be sure that statistics will be produced which won't compromise security be released?

If the answer to each question considered separately is 'yes', then 'safe use' of the data should be achievable.

The Anonymisation Decision Making Framework (Elliot at al) takes a similar approach: making a step-by-step decision thinking about how risky data access can be, given a set of controls that may or may not be in place. Ultimately, if all the right parameters are in place, there should be no reason to deny access: the risk is sensibly managed.

Conclusion

While legal and technical frameworks exist for enabling safe access to data, often IAOs, as the people charged with making day-to-day decisions about data access, can be reticent about allowing novel sources of data to be accessed for research. Their perception about whether use of the data has a sufficient legal basis may be based on misguidance about the benefits to society of the research; and when coupled with an attitude to risk weighted firmly towards aversion, decisions on data access tend to be turned down or delayed. The legal frameworks available are sometimes not applied in the spirit in which they were written.

Because of this, the 'default-close' position tends to be the one reached, and so the benefits of data access are not realised. This is despite a plethora of guidance available which could enable balanced, legal decisions to be made without much delay. Without a change to the profile of an IAO, access to novel data sources will continue to be slow, with confidentiality risk being cited as the cause, often unnecessarily.

This problem will be compounded if a risk/reward framework is biased towards risk-averseness and confidentiality protection: few civil servants are rewarded for enabling access to data.

References

Desai T., Ritchie F., and Welpton R. (2016) *The Five Safes: designing data access for research.* Working papers in Economics no. 1601, University of the West of England, Bristol. January

Ritchie F. (2016) "Can a change in attitudes improve effective access to administrative data for research?", Working papers in economics no. 1607. University of the West of England, Bristol.

Elliot, Mackey, O'Hara, Tudor (2016) "The Anonymisation Decision Making Framework", available at https://ukanon.net/wp-content/uploads/2015/05/The-Anonymisation-Decision-making-Framework.pdf