

Faculty of Business and Law

User-focused threat identification for anonymised microdata

Hans-Peter Hafner¹, Felix Ritchie² and Rainer Lenz³

¹ Saarland State University of Applied Sciences ² University of the West of England, Bristol ³ Technical University of Dortmund

Economics Working Paper Series 1503



User-focused threat identification for anonymised microdata

Hans-Peter Hafner, Felix Ritchie and Rainer Lenz

Abstract When producing anonymised microdata for research, national statistics institutes (NSIs) identify a number of 'risk scenarios' of how intruders might seek to attack a confidential dataset. This paper argues that the strategy used to identify confidentiality protection measures can be seriously misguided, mainly since scenarios focus on data protection without sufficient reference to other aspects of data. This paper brings together a number of findings to see how the above problem can be addressed in a practical context. Using as an example the creation of a scientific use file, the paper demonstrates that an alternative perspective can have dramatically different outcomes.

Key words: statistical disclosure control, data protection, microdata anonymisation, big data

JEL codes: C81, C18

Address for correspondence: Dr Felix Ritchie, University of the West of England, Bristol, BS16 1QY; email: felix.ritchie@uwe.ac.uk

Acknowledgements and disclaimer This work draws on an independent research project funded by Eurostat. This paper was written after project completion, and no inferences should be drawn about Eurostat's policies or views on the matters discussed. All opinions expressed are those of the authors.

1. Introduction

One of the key functions of national statistics institutes (NSIs) is to produce research datasets from the same sources used for aggregate statistics. Allowing access to this microdata effectively leverages the NSI's investment in data collection. As data collected by NSIs are typically confidential, the dataset is rarely released 'as is' but has confidentiality protection measures applied to it.

All NSIs carry out this same function to a greater or lesser degree and have sponsored much research into the production of 'safe' datasets. Hence there is a large academic literature to support such processes, as well as automatic tools such as μ -Argus¹ and SDCMicro².

However, there is also a strong perspective about the way that the tools should be used. As several authors have noted, NSIs tend to be risk-averse, more comfortable with the 'policing' than the 'sharing' approach to data access and focused on the statistical product rather than the use to which it is put. This leads to best practice models that emphasise the protection of data even in the most extreme circumstances.

This paper argues that the strategy used by NSIs to identify confidentiality protection measures is seriously misguided. It leads to an excessively conservative approach which is not supported by evidence or required by law, and which can frustrate users. The

http://neon.vb.cbs.nl/casc/mu.htm

http://cran.r-project.org/web/packages/sdcMicro/index.html

consideration of extremes over-protects the data, and the focus on data protection per se ignores the opportunities for protection via other methods.

In recent years there has been a small but growing literature challenging the predominant view of risk assessment, both on theoretical grounds and by empirical evidence from forty years of dataset protection. The oft-cited argument that exceptional protection is enshrined in law seems dubious; most countries only require 'reasonable' protection. The implicit ethical stance of most literature ('only release if safe') has no more inherent validity than the alternative ('release unless unsafe'), and yet this has a demonstrable impact on decisions taken. Perhaps most importantly, there is a wealth of information about how researchers use such data files and this is rarely considered.

Changing the perspective can have significant consequences for the users and owners of the anonymised dataset. An evidence-based and user-centred approach to anonymisation takes into account factors other than inherent risk in the data. This can make substantial improvements to the utility of the dataset while preserving the low-risk nature of the data.

We bring together a number of findings to see how the above problems can be addressed in a practical context. Using as an example the recent creation by the authors of a 'scientific use file' (SUF) from multinational business survey data, the paper demonstrates that an alternative perspective can have dramatically different outcomes: in this case, from 100% perturbation of all continuous variables to perturbation of under 1% of the observations for just one variable.

The next section summarises the traditional approach. Section 3 then provides a critique and proposes an alternative strategy. Section 4 describes how this alternative model was used in the recent creation of an SUF from confidential business microdata, and the impact this had. Section 5 discusses the lessons learned and considers the implications of this more empirical approach on the growth in administrative data sources.

2. Common approaches to anonymisation

The ESSNet Handbook on Statistical Disclosure Control (SDC: Hundepool et al, 2010) provides a comprehensive overview on the discursive microdata anonymisation process. This Handbook, sponsored by Eurostat, contains a wealth of guidance on methods and tools, plus references for further information. While the Handbook has no statutory authority, it does provide a wide-ranging summary of received opinion at the time of writing, and is therefore used to reference this section.

The Handbook notes that microdata protection should be based upon knowledge of the use of the data, the access requirements, the potential for an intruder to match external datasets, and the structure of the data itself. Risk scenarios are based upon both spontaneous recognition and actively searching for an individual, possibly using record linkage; for detailed empirical studies see Lenz (2006). For both of these, it is possible to generate estimates of the likelihood of re-identification of an individual. These probabilities can then be used to compare alternative data protection methods.

Whilst changes in the probability of detection can be described in a straight-forward manner, changes in the utility of the data are harder to quantify, as this depends upon the likely uses of the data. Sophisticated analyses on the effect of anonymisation to the analytical validity of microdata can be found in Ronning et al. (2005).

3. Critique of common perspective

There are three major concerns about the way microdata protection is implemented in practice. Two can be seen as failures to use evidence; the third is a case of failure of the theoretical framework for decision-making.

a. Focus on data protection

Microdata sets are described as 'public use files' (PUFs, available without restriction), 'scientific use files' (SUFs, available to a restricted set of researchers), or 'controlled access files' (CAFs, available only within an environment controlled by the NSI), but it is questionable how much attention is paid to these different surrounding conditions.

In particular, there is always the assumption that these files are subject to an intruder threat. For CAFs and SUFs this is a difficult case to make. Good practice requires the removal of direct identifiers (names etc.) from such files, so identification is only possible indirectly, implying some effort on the part of the researcher. For CAFs this effort is monitored. For SUFs, the effort is not monitored but effort is still required (see Lenz, 2006); in both cases it could be argued that, if intruder threat is a genuine risk, the problem lies with accreditation procedures and not with anonymisation.

Such evidence as there is suggests that intruder modelling is unrealistic. There are no cases (to the authors' knowledge) of *malicious* misuse of CAFs or SUFs in the ways identified by standard risk scenarios. There is ample evidence of SUF/CAF researchers making *mistakes*, or CAF researchers deliberately *circumventing procedures* to make life easier for themselves – but not to deliberately de-anonymise the data. Moreover, even such non-malicious outcomes are rare. One author's ten years of managing CAF usage saw three deliberate acts of misuse and another ten or so genuine mistakes, set in the context of some ten thousand user visits. Within this, the deliberate misuses were all the result of researchers trying to reorganise processes for their own convenience, not to re-identify data. The authors' experience with managers of SUF and CAF releases across the world suggests that this outcome is expected³.

In summary, for SUFs and CAFs empirical evidence suggests that factors other than protection of the data dominate the likelihood of successful protection; such non-data control measures have a forty year record of demonstrable effectiveness.

For PUFs, it could be argued that intruder threat is a genuine risk, as potentially it only needs one person in the world to have sufficient malice or prurience to try to breach confidentiality protection. However, this assumes that the data are sufficiently interesting to make the effort worthwhile. For example, if John Smith works at a local bakery, this could be determined from his Labour Force Survey responses; but it may be easier to find it out by using the internet, social networks or watching him walk to work. Yes, the Labour Force Survey data is confidential and should be protected – but is it of sufficient value to make an attack worthwhile⁴?

It could be argued that no NSI would willingly share information on a deliberate breach because of the poor publicity, but the relatively small size of the international data protection community militates against such as case, particularly when the errors that do occur are freely discussed amongst the community. One could

true but not practically helpful.

also argue that successful malicious breaches have occurred but remain undiscovered, which is theoretically

Of course the LFS does hold much more sensitive data such as health conditions. John Smith's occupation is an example of a low-value characteristic, such as those included in the 2002 Teaching version of the UK Labour Force Survey.

b. Worst-case scenarios and spurious objectivity

As noted above, it may be easier to get agreement on worst-case scenarios, as they seem easier to define. This makes sense in the context of academic research articles, where the aim is to compare methods using a common base wherever possible. Such assumptions can then allow *relative* effectiveness of methods to be assessed, which is important in developing understanding of the impact of different techniques.

It does not however follow that worst-case scenarios should be used in planning in practice. First, there is no evidence to suggest that typical worst-case scenarios ever manifest themselves. It is well known that there are large differences between data from official statistics and external commercial databases (eg. Lenz, 2006; Hafner, 2008). This implies a kind of extra protection for the data that adds to the protection achieved by anonymisation procedures. However the NSI often conduct a worst case scenario as they match the anonymized data to the original survey data. One reason for this practice is that it can be very time consuming and expensive to generate a realistic external data source for every survey since the commercial data have to be purchased from database providers and the identifiers often have to be harmonized manually. But the result of a non-realistic worst case scenario may not be treated the same way as the result of a realistic scenario. This is the fault in the line of argument the NSI commits, and so they prevent the release of more and better microdata for the scientific community. In addition a researcher who is sufficiently incentivised to undertake the risk and effort of matching is likely to have a number of different possibilities available (including the more simple one of talking to unsuspecting NSI staff).

Second, these assumptions are defended on the grounds that all practical measures need to be taken to protect the data. This is unlikely to be true. All statistical legislation leaves the level of protection as something to be determined in specific context; increasingly, laws explicitly state that only 'reasonable' protection need be provided (for example, Eurostat⁵ or the German construct of De Facto Anonymity⁶). The law in effect is recognising that worst-case scenario planning is unlikely to be good for society: designing strategy based on extreme hypothetical outcomes imposes costs on society which a more reasonable view of the likelihood of events would avoid.

Finally, even the objectivity of worst-case planning is suspect. Skinner (2012) argues that protestations of 'objectivity' in risk assessment are misleading; the framing of the risk assessment is decided by the NSI on subjective criteria. For example, the ESSNet Handbook describes a potential 'conservative and worst case scenario' with only one known external data source being used for matching and with design, but not response, weights available. Clearly, both assumptions are debatable, and an NSI adopting these assumptions is making a subjective decision.

Once it is recognised that worst-case scenarios are (a) inefficient (b) not supported by evidence (c) not required by legislation and (d) as subjective as any other, their use in decision-making comes into question.

c. The default position

The default perspective of most NSIs is defensive: no data can be released unless it can be shown to be 'safe'. However, the NSI could take the co-operative perspective that all data

Article 3 "Definitions" of the Commission Regulation (EC) No 223/2009

⁶ German Federal Statistics Law §16(6)

will be released unless it presents a disclosure risk⁷. Ritchie (2014a) demonstrates how these two perspectives will generate different outcomes, with the former almost certain to restrict data access much more. Typically this arises because NSIs have insufficient user input to influence the discussion and overcome security concerns (the 'diffuse benefit and concentrated cost' often associated with lack of government action; Moore, 2010; Ritchie, 2014b).

This defensive perspective is reflected in the lack of discussion in meetings and the literature. At the 2013 UNECE meeting on statistical confidentiality, two sessions on data access were organised. With the honourable exception of the Italian NSI ISTAT, no papers analysed user needs; all other papers explained how they were 'opening up' data access (that is, the default is 'no release') and this should be seen as a bonus for users.

d. Summary

The standard approach to anonymisation suffers from three failings: a focus on data protection as the prime guarantee of confidentiality; worst-case planning; and an approach to confidentiality driven by the need to avoid failure rather than maximise public benefit. Underlying each of these problems is an approach to risk that suggest is focuses on theory rather than evidence and data rather than environment. Hence we describe the 'typical' approach as theory-based and data-centric.

In contrast, we propose a model which is user-centred and evidence based. The following section describes the implementation of such an approach.

4. Example of an evidence-based risk assessment: the 2010 CIS⁸

The Community Innovation Survey (CIS) is a business survey carried out in all EU countries. Eurostat distributes a subset of country files, anonymised as 'scientific use files' for research purposes. However, uses have been very small and the perception exists that the existing anonymisation method has created, at best, a teaching dataset.

In 2013 Eurostat commissioned a review of the protection strategy to create the 2010 CIS SUFs with recommendations for changes, if any. Whilst the review recommended a number of significant changes, we focus on the risk scenario and the consequences for protection mechanisms.

e. Identify user needs

The study analysed 11 research papers using the CIS in SUF- and CAF-form in different countries (official documents from NSIs or other government departments were also analysed but these consisted exclusively of simple tabulations). In addition, the authors could draw on nine years' observations of researchers using earlier versions of the CIS in a controlled environment. Finally, a Google Scholar search was carried out. These confirmed that the overwhelming use of the CIS by researchers (as opposed to government agencies who hold the source data) was marginal analysis, particularly linear and non-linear regression.

This section is a summary of Hafner et al (2014). Specific references are not given as the form of publication is still being decided by Eurostat.

Many organisations formally support the 'co-operative' approach, but this does not necessarily happen on the ground. One author regularly addresses groups of data professionals; in shows of hands, respondents typically agree that the co-operative perspective is prefereable, but that the defensive perspective is their organisation's normal position.

f. Identify the user environment and risks

There is no evidence of malicious use of SUFs by genuine researchers. There is evidence of accidental and deliberate misuse which has the consequence of breaching confidentiality rules or procedures (Desai and Ritchie, 2009).

For business data, the most significant risk is spontaneous recognition of outliers. The researcher may publicly speculate on the identity of the firm. Alternatively the researcher may try to augment or compare the SUF with data from external sources. Note that it is <u>not</u> a risk that a researcher spontaneously notes the characteristics of an observation and muses on the company identity but does not follow up - there has been no disclosure to an unauthorised person, and no deliberate attempt to identify a company.

In general, outputs from genuine research are low risk, but there are a large number of categorical variables in the CIS and the interest in them makes the potential for disclosure by differencing larger. There is also the risk of group disclosure. The categorical variables in the CIS make saturated or empty cells more likely: for example, there may be many cells in a table where all companies undertake a specific form of innovation. However, as identified above, the main research interest in CIS use is in marginal analyses, where the disclosure potential is negligible.

There is always the risk from a misperceived output; for example, a naïve reader of a paper could assume that a statistic refers to a single company even if it does not. In this case, the risk is not to confidentiality but to the reputation of the organisations collecting and distributing the data.

g. Evaluate risks

Re-identification risk arises from publicly available classification data (company size, head office location etc) and from extreme values in continuous attributes, such as very high turnover. However, practical experiments done by the authors and others in this field (for example, Hafner, 2008; Bauer, et al. 2009) suggest that exact matching on continuous variables is not a practical concern, although a broad search on industrial classification and location might be more effective. In addition, identification may also arise from matching to external databases, but the sampling frame is the Eurostat-compliant business register, which is designed to be a *statistically* accurate reflection of economic activity, not *financially*. As is well-known (eg. Evans and Ritchie 2006), NSI business registers are difficult to reconcile with publicly available accounting information which makes extensive use of financial engineering.

These two factors provide considerable uncertainty about which companies are included in the data, and in which organisational form. Finally, much of the data in the CIS, while useful for research, has a low disclosure value. For example, being able to identify that Company X has engaged in product innovation over the period 2008-2010 is, technically, a breach of information supplied in confidence; but it is of negligible commercial value, and so unlikely to be a target for hackers.

In summary, spontaneous recognition is feasible but unlikely to have sufficient certainty to be worthwhile; a successful and informative match is theoretically possible but the practical problems are large. Most importantly, matching requires the researcher to actively search for the company; it is not an outcome of spontaneous recognition. The SUF licence agreement forbids attempting to identify any respondent; evidence suggests this is credible. Therefore, it appears that the risks of deliberate disclosure associated with researcher inquisitiveness are of a very low order.

h. Identify residual risks

This led to three credible risk scenarios:

- I) A researcher publishes a magnitude table with one or two observations in a cell
- II) A researcher comments on the dominance of one unit in a cell
- III) A researcher comments on the dominance of one unit in the dataset

These are all expected to arise as a result of error on the part of the researcher.

5. Impact

As noted above, the project recommended a wide range of changes to the anonymisation strategy, not all of which are relevant here. However, the way the risk scenarios were defined had important implications for the protection strategy.

The previous anonymisation strategy stated that deliberate misuse was not deemed to be a risk. However, the logic of this position was not followed through: no other explicit risk was identified; nevertheless, all observations were deemed to be potentially problematic. The problems were to be addressed by microaggregation of all metric variables and global recoding. This was also argued to reduce the need to test for and address dominance problems. The conceptual framework used in that case was defensive: 'apply protection until it is safe to release'.

In contrast, the revised risk scenario implied a very different protection model. As the key risk was identified as accidental disclosure, only measures to tackle dominance and small cell count were put in place, apart from a global recode of employment. In effect, only observations at risk were perturbed; the conceptual framework was 'apply protection only if it is demonstrably necessary; otherwise release'. Moreover, in deciding observations at risk, the team took account of the known disparity between published and surveyed employment data to provide additional arguments why the data was inherently safe.

The impact of this was to reduce the microaggregation from 100% of records to less than 1% for all countries; some countries saw no change to their data at all. Some sample linear and non-linear regressions were run, demonstrating considerable smaller impacts on coefficients. However, the strategy was also able to tackle dominance problems omitted in previous years. The 'do not disturb' strategy of Ichim (2007) and Ichim and Franconi (2008) found similar results.

6. Discussion and conclusion

Theoretically, microdata protection is a well-established mature field with a great deal of advice for NSIs trying to make data available. However, in practice, there is a concern that data protection takes precedence over the user experience. This has been a complaint from users for years but in recent times the data protection community has begun to question the profoundly conservative outlook found in NSIs.

The example discussed, of creating an SUF for the CIS, shows that a change in attitude can have significant consequences. No new methods were developed: protection was a combination of recoding and microaggregation. The difference came in the default perspective of the research team; the use of evidence in assessing disclosure risk; a realistic interpretation of what counted as 'reasonable' protection; and an explicit allowance for protection measures in the access environment. The end result was a dataset with more protection but much less impact on users.

Future trends in data are moving away from surveys to administrative data sources. These will present additional problems. For example, PUFs based on administrative data will be accessible to the office workers who have access to the original data. This implies that the problem of matching to external databases will become much more prevalent. Given the amount of perturbation needed to protect against matches in an identical data source, it may be worthwhile beginning to gather more evidence on the effectiveness of non-data-based protection mechanisms.

References

- Bauer, T., Bachteler, T., Bender, S., Huber, M., Schnell, R., Dundler, A., and Engel, D. (2009) Why are firms treated in a different way than individuals? Results of a reidentification experiment with real data. Discussion paper. Ruhr-University Bochum
- Desai, T. and Ritchie, F. (2010) Effective researcher management. Work session on statistical data confidentiality, Bilbao 17-19 December 2009
- Evans, P. and Ritchie, F. (2009) UK Company Statistics Reconciliation Project: final report, Report for the Department of Business Enterprise and Regulatory Reform; URN 09/599
- Franconi, L. and Ichim, D. (2008) Community Innovation Survey: comparable dissemination. In: Work session on statistical data confidentiality, Manchester 2-4 December 2007, Eurostat Methodologies and Working papers; pp11-23
- Hafner, H.-P. (2008) Die Qualität der Angriffsdatenbank für die Matchingexperimente mit den Daten des KSE-Panels 1999 2002. Mimeo. IAB
- Hafner, H.-P., Lenz, R., Ritchie, F. (2014): Report for Eurostat on the creation of CIS scientific use files. Eurostat, June.
- Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Lenz, R., Longhurst, J., Schulte Nord-holt, E., Seri, G. and De Wolf, P. (1010) Handbook on Statistical Disclosure Control, ESSNet SDC http://neon.vb.cbs.nl/casc/.\SDC Handbook.pdf
- Ichim, D. (2007) Microdata Anonymisation of the Community Innovation Survey Data: A Density Based Clustering Approach for Risk Assessment. Dokumenti Istat 2 http://www3.istat.it/dati/pubbsci/documenti/Documenti/doc_2007/2007_2.pdf
- Lenz R. (2006) Measuring the disclosure protection of micro aggregated business microdata an analysis taking as an example the German Structure of Costs Survey, Journal of Official Statistics **22** (4), 681-710
- Moore M.H. (2010) Break-Through Innovations and Continuous Improvement: Two Different Models of Innovative Processes in the Public Sector. Public Money and Management January v44
- Ritchie F. (2014a) Access to sensitive data: satisfying objectives rather than constraints, J. Official Stat., October
- Ritchie F. (2014b) Resistance to change in government: risk, inertia and incentives. Working papers in Economics no. 1412, University of the West of England, Bristol. December
- Ronning, G., Sturm, R., Hoehne, J., Lenz, R., Rosemann, M., Scheffler, M., and Vorgrimler, D. (205) Handbuch zur Anonymisierung wirtschaftsstatistischer Mikrodaten. Statistik und Wissenschaft, Band 4, DeStatis
- Skinner, C. J. (2012) Statistical disclosure risk: seperating potential and harm. International Statistical Review **80** (3), 349-368

Recent UWE Economics Papers
See http://www1.uwe.ac.uk/bl/research/bristoleconomics/research for a full list

<u> 2015</u>	
1503	User-focused threat identification for anonymised microdata Hans-Peter Hafner ¹ , Felix Ritchie ² and Rainer Lenz ³
1502	Reflections on the one-minute paper Damian Whittard
1501	Principles- versus rules-based output statistical disclosure control in remote access environments Felix Ritchie and Mark Elliot
<u> 2014</u>	
1413	Addressing the human factor in data access: incentive compatibility, legitimacy and cost- effectiveness in public data resources Felix Ritchie and Richard Welpton
1412	Resistance to change in government: risk, inertia and incentives Felix Ritchie
1411	Emigration, remittances and corruption experience of those staying behind Artjoms Ivlevs and Roswitha M. King
1410	Operationalising 'safe statistics': the case of linear regression Felix Ritchie
1409	Is temporary employment a cause or consequence of poor mental health? Chris Dawson, Michail Veliziotis, Gail Pacheco and Don J Webber
1408	Regional productivity in a multi-speed Europe Don J. Webber, Min Hua Jen and Eoin O'Leary
1407	Assimilation of the migrant work ethic Chris Dawson, Michail Veliziotis, Benjamin Hopkins
1406	Empirical evidence on the use of the FLQ formula for regionalizing national input-output tables: the case of the Province of Córdoba, Argentina Anthony T. Flegg, Leonardo J. Mastronardi and Carlos A. Romero
1405	Can the one minute paper breathe life back into the economics lecture? Damian Whittard
1404	The role of social norms in incentivising energy reduction in organisations Peter Bradley, Matthew Leach and Shane Fudge
1403	How do knowledge brokers work? The case of WERS Hilary Drew, Felix Ritchie and Anna King
1402	Happy moves? Assessing the impact of subjective well-being on the emigration decision Artjoms Ivlevs
1401	Communist party membership and bribe paying in transitional economies Timothy Hinks and Artjoms Ivlevs

<u>2013</u>	
1315	Global economic crisis and corruption experience: Evidence from transition economies Artjoms Ivlevs and Timothy Hinks
1314	A two-state Markov-switching distinctive conditional variance application for tanker freight returns Wessam Abouarghoub, Iris Biefang-Frisancho Mariscal and Peter Howells
1313	Measuring the level of risk exposure in tanker shipping freight markets Wessam Abouarghoub and Iris Biefang-Frisancho Mariscal
1312	Modelling the sectoral allocation of labour in open economy models Laura Povoledo
1311	The US Fed and the Bank of England: ownership, structure and 'independence' Peter Howells
1310	Cross-hauling and regional input-output tables: the case of the province of Hubei, China Anthony T. Flegg, Yongming Huang and Timo Tohmo
1309	Temporary employment, job satisfaction and subjective well-being Chris Dawson and Michail Veliziotis
1308	Risk taking and monetary policy before the crisis: the case of Germany Iris Biefang-Frisancho Mariscal
1307	What determines students' choices of elective modules? Mary R Hedges, Gail A Pacheco and Don J Webber
1306	How should economics curricula be evaluated? Andrew Mearman
1305	Temporary employment and wellbeing: Selection or causal? Chris Dawson, Don J Webber and Ben Hopkins
1304	Trade unions and unpaid overtime in Britain Michail Veliziotis
1303	Why do students study economics? Andrew Mearman, Aspasia Papa and Don J. Webber
1302	Estimating regional input coefficients and multipliers: The use of the FLQ is not a gamble Anthony T. Flegg and Timo Tohmo
1301	Liquidity and credit risks in the UK's financial crisis: How QE changed the relationship Woon Wong, Iris Biefang-Frisancho Mariscal, Wanru Yao and Peter Howells
<u>2012</u>	
1221	The impact of the quality of the work environment on employees' intention to quit Ray Markey, Katherine Ravenswood and Don J. Webber
1220	The changing influence of culture on job satisfaction across Europe: 1981-2008 Gail Pacheco, De Wet van der Westhuizen and Don J. Webber
1219	Understanding student attendance in Business Schools: an exploratory study Andrew Mearman, Don J. Webber, Artjoms Ivlevs, Tanzila Rahman and Gail Pacheco