Brisbane, 16 Feb, 2018

Response to the Consultation Paper "Proposal to remake the Statistics Determination 1983"

This response is by Per Davidsson, Professor of Entrepreneurship and Director of the Australian Centre for Entrepreneurship Research (ACE) at QUT, Brisbane. I make the submission as an individual who has considerable experience from working with the type of data concerned for both scholarly and policy-oriented purposes.

1. The consultation paper suggests sound and very welcome changes of moderate proportions, where the starting point – in international comparison – is an extremely over-cautious stance towards data safety based on severe exaggeration of the risks of data access and under-estimation of the benefits foregone associated with restricted researcher access to data.

2. A personal testimony: During the 1990s, in close collaboration with Statistics Sweden on the input (data) side and a major Government agency (NUTEK) on the output side, I engaged in three major projects based on secondary business data. These projects yielded unique new knowledge reported in 10 scholarly journal articles (including one with 1,000+ citations); 10 book chapters, and 10 policy-reports (as well as at events and face-to-face) which significantly informed concurrent policy debates about the role of start-ups and firm growth in job creation and regional development. The access to current, own-country data fostered a natural combination rather than separation of scholarly and policy-oriented pursuits as well as active involvement beyond the country borders through engagement with the OECD and the EU Commission, facilitating knowledge exchange. As an important by-product, our research led to a change in the US Bureau of Labor Statistics' reporting of job creation by firm size classes (www.bls.gov/opub/hom/bdm/pdf/bdm.pdf). By contrast, since my move to Australia in 2004 (citizen since 2007) I have published a total of two scholarly articles based on Australian business statistics. By the time of publication, these data (BLS) where more than 10 years old and of limited interest for concurrent policy. We did not work with the then more current Business Longitudinal Databases (BLD) because access to and working with the data was simply so hard that it was not worth the effort.

3. Likewise, despite very active efforts we have been unable to contribute new, policy-relevant knowledge based on the BLADE (formerly EABLD) data base. I and the centre I lead (ACE) have been able to develop a fruitful collaboration with DIIS around non-ABS/ATO data sets (CAUSEE [eprints.qut.edu.au/49327/] and Global Entrepreneurship Monitor [gemconsortium.org/data]. As these projects' ability to yield novel insights started to wane, I had hoped to continue a fruitful collaboration based on BLADE. Specifically, during my work in the 1990s, in collaboration with Statistics Sweden we pioneered a method for distinguishing between organic (internal) and acquisition-based growth of firms. This distinction is of scholarly interest, but also crucially important for policy interests in job creation (acquisition represents move of existing jobs, not the creation of new ones). I therefore engaged with ABS and DIIS in the development of BLADE; got assurances that this distinction was possible to make, and hired a PhD student for the purpose. To this date we do not have data access, and instead of contributing also to policy-relevant knowledge in Australia, the student is basing her dissertation work on data from Sweden and Germany. This is probably but one of many examples of benefits foregone through overly restrictive data access criteria.

4. The five safes. Of course, data safety is a very important issue. It is therefore important, as suggested, that "It will remain a criminal offence for a person to breach these conditions and may attract penalties including a fine, imprisonment, or both". However, the consultation paper (uncritically?) portrays the 'five safes' as 'best practice'. A best practice ought to consider both costs and benefits, and it is not clear that the 'five safes' – as conceived or as implemented – does this in a considered and systematic fashion. A country could reduce the number of road causalities to (close to zero) by introducing and enforcing a 10 km/h universal speed limit and proclaim it as 'best practice', but it would rule out the vast majority of safe driving that can occur at considerably higher speed and have detrimental effects on the economy of the country and the well-being of its people.

5. Can researchers be trusted? As a scholar, I have repeatedly been deeply disappointed by reports of colleagues who in their misdirected career ambitions have resorted to data fabrication and other types of research fraud. However, *I have never, ever heard of a case of a researcher either gaining financial/other private benefit or causing harm to a business included in a data set by identifying the case and acting on the obtained information.* Do such cases even exist, and what are the numbers? It appears to me that much of the safety concerns focus on an imaginary problem rather than on the real problems associated with untrustworthy researchers.

6. Suggestion: For researchers to use the data, it is important that they can do so on their own computer. Requiring that they do so in a "safe [physical] environment" is impractical and will probably reduce usage by 90+ percent. It is also important that variable values are not change so that the data produce false results. For most research purposes, name or ABN numbers are not needed; very detailed industry and location information (for example) is not needed, either – broader classifications that do not allow identification even if combined will do. By and large, researchers have no interest whatsoever in identifying the cases in the data set. When identification is needed to link data from different sources it is acceptable to do so via an intermediary if an efficient routine therefore is in place. If parts of the data set cannot be released (e.g. cases identifiable by their size or other uniqueness; particularly sensitive variables) then release a partial data set and let those who really need these parts of the data either "order" a particular analysis after "playing" with the partial data set, or do the full data analysis themselves in the "safe [physical] environment" after they have been able to undertake preliminary analyses on the released, partial data set.

In all, the consultation paper suggests some long overdue but very welcome changes that can significantly improve Australian scholars' contributions to practice. We live in an international world and one where academics are under pressure to (firstly) contribute to their scholarly discipline, so without access to high-quality, Australian data they will collaborate with colleagues with access to such data from other countries. It is in the Australian tax-payer's interest that we devote more time to data that can simultaneously satisfy scholarly goals and Australia-specific, practical purposes.

Sincerely,

Per Davidsson

http://staff.qut.edu.au/staff/davidssp/