What are official statistics for anyway?

Steve MacFeely

Central Statistics Office, Cork, Ireland
Centre for Policy Studies, University College Cork, Cork, Ireland

Contact: steve.macfeely@cso.ie

Abstract:

In Europe and in most developed countries a vast amount of official statistics are compiled; short-term and structural data covering a broad spectrum of economic, environmental and social activities. The collection and compilation of these data are costly, both to businesses and the public who indirectly pay for the collection through taxes and directly through the time and administrative costs or burden of supplying the data. As recompense, most nations view official statistics as a public good, central to the accountability and transparency necessary for a functioning democracy and disseminate them free of charge.

Yet beyond the halls of academia and policy making these tremendous resources appear to be under-utilised. Why is this and does it matter? For some, official statistics generally and business statistics in particular, are simply part of the supply chain that supports the compilation of the national account and this is justification enough. But there is a growing body of evidence to suggest there is a problem; a problem that has the potential to adversely affect the quality of the data being compiled.

This paper outlines some recent research highlighting the link between data quality and perceived usefulness. A number of examples of where data might be useful to business are also illustrated.

Keywords: public good, big data, data quality, perception

Introduction

In Europe and in most developed countries a vast amount of official statistics are compiled; short-term and structural data covering a broad spectrum of economic, environmental and social activities. The collection and compilation of these data are costly, both to businesses and the public who indirectly pay for the collection through taxes and directly through the time and administrative costs or burden of supplying the data. As recompense, most nations view official statistics as a public good, central to the accountability and transparency necessary for a functioning democracy and disseminate them free of charge.

Yet beyond the halls of academia and policy making these tremendous resources appear to be under-utilised. Why is this and does it matter? For some, official statistics generally and business statistics in particular, are simply part of the supply chain that supports the compilation of the national account and this is justification enough. But there is a

growing body of evidence to suggest there is a problem; a problem with the potential to adversely affect the quality of the data being compiled.

This paper outlines some recent research highlighting the link between data quality and perceived usefulness. A number of examples of where data might be useful to business are also illustrated.

A changing landscape

The landscape in which official statistics operate is changing and changing fast. In a world where our increasing day-to-day dependence on technology, social media and electronic transactions are leaving significant 'digital footprints' in our wake and where the business model for many electronic services generates revenue from advertising rather than the provision of a core service, the monopoly held by official statistics for so long, to provide, timely, good quality statistics for free, is being challenged. The torrent of so called 'Big Data'[1] now generated as a by-product of these new digital services has been described as a '*Data Deluge*' (Vale, 2012). In fact the volumes of data being generated now are so great and increasing in volume at an exponential rate, that half of the data that exist today were only generated in the past two years (Shatter, 2013). According to Keane (2013, p.1) "*Stored digital content is doubling two years, reaching one zettabyte*"[2] in 2012. Many modern service providers are now disseminating statistics, which to the untrained or undiscerning eye resemble official statistics, in order to attract media attention. Increasingly, official statistics are being forced to differentiate themselves from these new 'by-product' statistics by highlighting their relative quality which comes from conceptual rigour, methodological quality and transparency of metadata. Competition always presents challenges and often times results in a better service, but in a situation where official statistics must adhere to increasingly rigorous quality standards with all the inherent trade-offs between quality and timeliness, whereas 'marketing' statistics do not necessarily adhere to any known standards, it is not clear what the future holds for official statistics.

A changing role

Today an indiscriminate and often times sensationalist media are presenting a vast array of data and analyses of varying quality to a public and business community who are insufficiently statistically literate to distinguish between quality standards or able to distinguish between fact and comment. In a world, where perhaps there are too many data (there is without question too much comment being portrayed as data or fact), potential users are increasingly likely to be confused and may select inappropriate data upon which to inform themselves or base their decisions. As societies and economies are becoming more globalised and complex, official statistics, in order to properly measure social and economic transactions are unavoidably becoming more complex in parallel. Unfortunately, in most cases, to derive value from official statistics, the user must invest

[1] The Wikipedia definition of big data is 'a collection of datasets so large and complex that it becomes it becomes difficult to process using on-hand database management tools or traditional data processing applications.

[2] A Zettabyte is a billion billion megabytes

some time informing and educating themselves; an investment many appear unwilling to make.

But users, already confused and bewildered by the massive volume of statistics available, and increasingly time poor, may find themselves being attracted to statistics that are visually attractive or appear simple or straightforward. Complex concepts or methodologies are rarely seen as attractive attributes by users.

Arguably then, the data deluge has created new or supplementary roles or responsibilities for national and international statistical institutes; to re-package our products to make them more user friendly without compromising their integral quality and to educate users so that they are better able to select and use appropriate data to suit their purposes. Thus brand building and education must become key objectives for NSIs in the future. In doing so, NSIs will provide a better service to their customers and protect their market position.

A 'trusted' brand

All statistics, including official statistics, are frequently misused. Hence the wonderful phrases and jokes about statistics, the most famous of which is attributed to Mark Twain – '*Lies, damned lies, and statistics'*. These quips are humorous because we all know they contain more than a grain of truth.

Determining whether such misuse is inadvertent or deliberate is difficult, and whatever the cause misuse occurs with frustrating regularity. Nor is it easy to address (it is not an enjoyable task accusing someone of being either stupid or dishonest or both). The end result is often a misinformed or confused public. This situation is less than satisfactory and raises difficult and challenging questions for NSIs – Are traditional official statistics fit-for-purpose? If they are not being used, or being continually being abused does it makes sense to produce them? Too often it is easy to blame the customer, often more-or-less statistically illiterate, but perhaps we should look closer to home.

To many, statistical results are not clear or intuitive. In fact, to many, the results seem downright improbable, impenetrable, unrealistic or just downright crazy. In many cases this confusion arises because users don't read the metadata, the background notes or the small print. So, too often, users think they understand the data and their implications when in fact they don't (for an example of this, compare how a typical householder and a national accountant might define or describe 'household savings'). But in other cases NSIs must accept that the message is not always as clear as it could be. No doubt this contributes to the perception that statistics are less than useful.

Nevertheless, despite these challenges, official statistics retain some major strengths. Not least, official statistics are impartial and politically independent. The importance of these attributes should not be underestimated but rather highlighted at every opportunity. Official statistics are compiled without fear or favour, independent of ideology or vested interest.

Official statistics must ensure the quality of the data being produced is robust and remains fit-for-purpose. Furthermore it is imperative that NSIs are able to prove that this is the case. Therefore official statisticians must be able to quantify the relative precision of their data. It is not enough to say the quality of data is fit-for-purpose, that quality standard must be defined and supporting evidence must be provided to demonstrate this standard

has been achieved. As costs escalate and intolerance to response burden grows it is critical that the lessons from emerging research on respondent motivation and the impact on quality is taken seriously. While this applies to all aspects of survey work, the recent work published by Giesen, Bavdaž and Bolko (2013) highlighting lessons for enterprise statistics must be taken seriously by NSIs.

What is the limit of official statistics?

While some official statistics are relatively straight forward and easy to understand, many aren't. Furthermore, several official statistics appear simple and straightforward but in fact are anything but; two examples, of significant policy relevance and high public interest but where the complexity is not fully appreciated are discussed below:

The first example is Residential Property Price Indices (RPPIs). These indices have jumped in public popularity in recent years, forming the basis of many dinner conversations. They have also assumed increased economic importance since the financial crash in 2008. In Ireland, there are several residential property price indices to choose from; compiled by banks, property agencies, auctioneers and of course the Central Statistics Office. While the official RPPI compiled by the CSO has a number of methodological weaknesses (most significantly the absence of cash based transactions) it is the only transaction based index and superior to the alternative indices available. Furthermore the methodology and metadata are clearly outlined and freely available (O'Hanlon, 2011). Meanwhile, the competitor indices are not based on any transaction data, but this critically important fact is ignored by commentators and users alike as their presentation is attractive (although often misleading). A particular flaw is the spurious level of precision and detail typically presented and accepted by unsuspecting users in good faith. Another weakness is the derivation of effectively meaningless statistics – 'the average price of a house' being the most popular. The user doesn't usually realise that most banks or estate agents are selling something and that a particular trend may suit their purpose. By contrast, NSIs are selling nothing, simply trying to inform users what is going on.

The second example is public-private earnings analysis. As the recession deepened in Ireland, and the private sector began to lose employment, the attention of the media and the public turned to the insulated status of public sector workers. During the early stages of the recession, official earnings were met with disbelief as counter-intuitively they didn't record any significant fall in hourly wage rates and ran contrary to anecdote. However subsequent studies showed that CEOs of private enterprises preferred to shed labour, or reduce hours rather than tackle the unit cost of labour; hence the dramatic loss of full-time labour (CSO 2010; Walsh 2012). The subtle distinction between take-home pay and hourly wages rates was lost on many. A more difficult proposition was how to explain the complexities of the public-private pay gap. Rather than publish a definitive number, the CSO presented a range, reflecting the lack of agreement in the literature as to best measure this complex subject (CSO, 2012). Several plausible models could justifiably the used and indeed a range of different parameters could be used within each multivariate model (Foley and O'Callaghan, 2010). Furthermore, the CSO also provided an analysis by percentile to illustrate that the pay gap is not uniform across the income distribution. These analyses were peer reviewed by international experts prior to publication. The main message that CSO wished to convey was that a simple, definitive answer did not exist and furthermore any attempt to give one would be overly simplistic

and misleading. While the report was welcomed by most economic commentators as balanced if slightly impenetrable, it was completely ignored by the media until one group of commentators launched a rather acerbic attack (Kelly et al, 2013) - then they had a story the public would understand.

These examples illustrate some of the challenges facing official statistics that can arise when trying to provide useful, accurate and balanced information. The public wants to know *the* answer. The media, reflecting this wants simplicity, even if that simplicity may yield misleading or meaningless results and commentary. Complex subjects are frequently interspersed with grey areas that defy simplicity; often understanding these grey areas is critical to comprehension. Unfortunately, the definitive answer is not always possible and attempts to produce one are often misleading and ideologically driven. The question for NSIs is whether this apparent trade-off, between impartiality and making statistics understandable is unavoidable.

What use are official statistics to business?

The use of statistics to academics, researchers and policy makers is reasonably well understood. What is perhaps less well understood and certainly under-utilised is how official statistics can be of use to businesses or enterprises. To the statistician the uses are obvious, so much so, the statistical community may have mistakenly made the implicit assumption those uses are as obvious to everyone else. The type of analyses that official statistics could contribute to are numerous. A number of examples have been outlined as illustrations:

- How am I doing vis-à-vis the market or competitors?
- What are my chances of survival if I establish a new business in sector x or in region y?
- What is the potential labour supply like – volume, education standard and age?
- Will ICT, or exporting, improve my productivity?
- Should I offshore – will it reduce my costs or improve my profit?
- Does size matter? Are there economies of scale – what are the tipping points?
- How are competitors reducing their wage bill?
- How are my customers doing? Is their income rising? Are their savings falling? Are they in negative equity? Do they have adequate pension provision?
- Where is unemployment highest? Or migration?
- What is the demographic profile of a local area? Age? Gender? Marital status? Nationality? Crime rate? Income levels?
- Who is exporting similar products or services? Where are they exporting to?
- What is the country or region importing? Are there opportunities to provide those goods or services locally?
- What might a typical supply chain for a sector look like?

Taking one of these illustrations as an example - an enterprise trying to identify business opportunities could start with national or regional Supply & Use (SUT) or Input-Output (IO) tables. These tables give a very good overview of a national or regional economy. In particular, they show what sectors are importing and exporting goods and services and the volume of that trade. These tables can be combined with merchandised trade and balance of payments data to better understand what commodities or services are being traded.

Thus gap or replacement analyses can be done to identify opportunities where international or inter-regional imports could be substituted with domestically produced goods or services. Reversing this process, enterprises can identify international export markets that are currently importing products or services that they are producing. Furthermore, IO tables also illustrate and quantify the supply chain for economic sectors, showing potential markets for their goods and services throughout the domestic economy.

Conclusion

Although there are a range of developments challenging the dominant position of official statistics, the justification or arguments in support of official statistics have never been stronger. The need for evidence-informed policy formulation and evaluation is increasingly understood and demanded by citizens. Equally, the understanding that democracy will be best supported by data compiled independent of ideology or political interference is more and more demanded. Furthermore, the public expect the quality of those data must be sufficiently robust to support the purposes to which they are being applied.

While this is intuitively understood, it can often be forgotten, or the trade-offs seem less serious when faced with superficially attractive alternatives. Thus it is critical that NSIs understand that official statistics must be marketed and promoted just like any other commodity. Counter-intuitively, they must be promoted more because they are free – frequently we learn the value of public goods are not properly understood or valued.

So NSIs must invest more thought and effort into how to promote and brand their product. The ABS has shown they realise this, some of their recent innovations are highly creative[3]. NSIs must promote the availability of data and improve the accessibility of those data; this means making it easier to find the data and also making to easier and more intuitive for non-specialist users to understand the data. To encourage businesses to respond and engage with statistics, tailored or bespoke analyses could be generated to demonstrate the usefulness of the data they are contributing to.

So in todays 'information age' NSIs must carefully consider a number of issues: their role in educating their users and potential users; how best to communicate the key messages from complex data and how to provide better context without compromising the appropriate limits of official statistics. NSIs should consider whether they need to quantify, not just the cost of statistics, but also the downstream benefits; some countries, for example UK and NZ have begun exercises to try and quantify the economic benefits of their large scale surveys.

[3] For example, their 'Run that Town' app – their interactive Census game.

References:

Giesen, Bavdaž and Bolko (2013), 'Blue-ETS'

CSO (2010), "Analysis of Wage Bill Change in Enterprises and Components of Change, Q3 2008 – Q3 2009 – Supplementary Analysis: Earnings and Labour Costs". www.cso.ie

CSO (2012), "National Employment Survey 2009 and 2010: Supplementary Analysis", www.cso.ie

Keane, M. (2013), "Big Data is Watching you", quoted in article by K. Lillington – Irish Times: Weekend Review – Saturday, April 27, 2013

Kelly, E., S. McGuiness & P. O'Connell (2013), "Comparing Public and Private Sector Pay in Ireland: Size Matters – Research note", *Quarterly Economic Commentary*, ESRI, Spring 2013.

Foley, P & F. O'Callaghan (2010), "Investigating the Public-Private Wage Gap in Ireland Using Data from the National Employment Survey 2007", Journal of the Statistical and Social Inquiry Society of Ireland, Vol.XXXIX, pp. 23-52

O'Hanlon, N (2011), "Constructing a national house price index for Ireland", *Statistical and Social Inquiry Society of Ireland*, Vol.40, pp. 167-196

Shatter, A, (2013), "…." *European Data Forum*, Dublin.

Vale, S, (2012), "…." *CSO Seminar on Administrative Data*, Dublin

Walsh, K (2012), "Wage bill change in Ireland during recession - how have employers reacted to the downturn", Statistical and Social Inquiry Society of Ireland, Vol.41, pp. 39-70