

COORDINATION OF DATA SOURCES

Willie A Maree

South African Reserve Bank

1. Introduction

We share a common interest in the compilation of *'Statistics that make sense'*. To this end data are sourced in various ways that are mostly complex and time consuming. Data are aggregated and translated into statistics. Statistics are purposefully packaged, contextualised and translated into information that is trusted to be meaningful. In this sense economic statistics attempt to provide with reasonable accuracy a prognosis of the health of the economy in a given measuring period. Statistics are assessed, enriched with more input from which we then derive other useful information such as direction and momentum. It is a culmination of activities that include measurement, assessment, often some debate and explanation. This brief exposition of the information creation process or information creation value chain over simplifies the intricate, expensive and time consuming activities of a typical economic research facility.

Towards creating *'Statistics that make sense'*, it is important to understand what is required to achieve just that. There is a combination of factors that will in the end determine value or nonsensical outcomes. It is a culmination of four significant components, i.e. data, implicit knowledge, technology and technological proficiency. It all begins with the sourcing and collection of data and statistics with us having more or less control over what is received. Mostly data are garnered by means of surveys. Without a solid (known) base, collecting data through surveys are often and rightly so perceived to be random.

While criticising is easy, everything is not as straightforward as it may seem. There is not a single silver bullet. When the offerings of modern technologies are considered, there is always a glimmer of hope. However, technology is not immune to problems and in terms of processing numbers, should not be allowed to create a false sense of security. Technology is only as good as its intended application, design and intelligent use. *'Closing the loopholes'* will continue to be a challenge. As endless as the potential of technology may seem, its value and limitations are directly related to the technological proficiency of the user. If your business is about information, this is a minimum requirement.

2. The Internet and Information

The 'Internet' is often misconstrued as a source of information. In the early years following its introduction, superior knowledge and "I got it off the Internet" was by novices often used as concluding arguments to lengthy debates. The internet is simply a medium similar to a newspaper, journal or tabloid that provides us with access to insights that may, or may not be useful. It is nothing more than a combination of connective technologies and services that enable the swift transmission of digitised stuff from one digital location to another. The word 'stuff' is deliberately chosen because unless assessed, nothing obtained from or accessed via the Internet must be taken for granted or used as authoritative information. Information obtained via the Internet still need to be subjected to the same scrutiny and pass the same stringent criteria that would normally apply to information otherwise obtained.

Following the advent of computerised systems and its potential, it was AT&T some fifty years ago that referred to the introduction to the *'information age'*. Some ten years following the introduction of the Internet, academics and significant role players realised that the technology created possibilities that transformed what was for very long hailed as the

'information age' to what is today quite appropriately known as the 'media age'. What is the difference then – one may ask? The emphasis has shifted to connectivity and social networking while terminology shifted from simply 'browser' to Google, Wiki-, Facebook, Twitter and YouTube, iPhone and iPad and 24/7. Adding to the complexity is a myriad of mega search engines, search engines and search portals - each boasting own niches, characteristics and search capability.

With rapidly advancing technologies, two fearsome concepts were introduced, i.e. 'Information overload' and 'Information anxiety'. 'Information overload' refers to *significantly more information the human brain needs to process or take into account to arrive at a specific decision*. 'Information anxiety' is *the difference between what people actually understood as opposed to what people think they should understand – somewhat simplified, the gap between data and knowledge*. The human cost of 'overload' is invoked when anxiety is caused by feelings of inadequacy. (Wurman, 1989, Information Anxiety, p 222). Typical symptoms of 'information anxiety' are the aimless seeking for more information and procrastination – endless delving into the information abyss until you get trapped and run out of time to complete assignments. Others refer to the 'Alice in Wonderland Syndrome'.

Escaping overload or anxiety is dependent on the ability of the knowledge worker to organise and plan activities better. It is imperative that the user must determine what types of information are most appropriate and adequate for the assignment and the preferred format. (Data, text, graphs, spreadsheets) The seasoned analyst or researcher is well aware of most likely reliable sources. Critical to the search and retrieval of data is to understand the difference between search engines and search directories. (The latter are much better catalogued, content better classified and indexed. Access however may require special permissions.) There are literally hundreds of search engines. It is important to learn about their strengths and niches, capability and limitations. It will save you much valuable time.

Distinguishing between clutter, noise and trusted information is equally important. The Internet is organised in domains which are usually identifiable by extensions such as .com, .org, .edu, .net, etc. Depending on the type of information that is required, it is expected that a paper published on a .org and .edu domain to be more trusted than a 'white paper' obtained from a .com. Learning and understanding the differences is imperative for the knowledge worker to master his/her field of expertise and be technological proficient.

3. Ownership and accountability

How important is ownership really? Referring to the strategic significance of information, according to Jacobs, key to the success of any organisation or business unit is *'information that makes sense.'* Information must therefore be current, accurate, relevant and understandable. (Jacobs, 1997, Real Time Strategic Change, p 198) The same applies to economic information. The subject dictates the type of information that is required. What should be determined is what data requirements are, how much of it, the desired reliability, potential sources, the ideal medium and format. Issues such as representation and coverage determine the relative richness and credibility of the information. Answers to the questions will point towards the technologies that are needed.

From the rather elaborate introduction we have learnt that it is imperative of the knowledge worker to have a somewhat more than a working understanding of available technologies. Some may be automated to be more use-friendly. To know what you know is a reassuring position for an analyst to be in. To know what you don't know provides an opportunity to source supplementary data/statistics towards *'Statistics that make sense'*. Not to know what you don't know is unsupported and a hopeless position to be in (shaky). Ensure that you understand the characteristics of your base.

In the early days of automation, business was forced to separate roles of data ownership, data management and input/output control. Data were sourced and managed on behalf of owners while control and lopsided workflow characteristics were incorporated in designs. Coincident with the introduction of the Internet, was an all new urgency and impatience. The Internet and associated technologies rendered such draconic processes obsolete. A brand new IT dictionary with new acronyms and altogether new jargon and concepts had to be introduced. Client portals can now be plugged into 'intelligent systems' that are 'rules based'. Data are now 'captured once at the source'. Where is the control? New intelligent systems shifts accountability and control back to ownership. The owner controls input, workflow, authentication, publication and access. Owners control data assets and everything of it. Separation of duties including data control and data management is regarded as undesirable.

4. Centralising sources of information

The centralisation of information sources is contentious. There has to be some agreement with regards to what is included and qualifies as information; metadata characteristics such as type become important, also characteristics regarding structured and unstructured qualities and the level of classification, indexing and tagging. It will determine the efficiency of retrieval and what technologies are more appropriate to use. It will significantly enhance detect and retrieval turnaround time. Sources need to be explored and assessed to determine which of the sources meet rigorous criteria to qualify as value propositions. Issues pertaining to intellectual property and terms of access need to be clarified. There should be a policy with regards to what is stored, what is linked and who manages the domain.

Another consideration is to determine if desirable, to what extent the development of a central repository of information accesses should combine resources with other areas in the Bank, particularly other economists. The Research Department already obtains some information from other areas in the Bank. Areas that are still unexplored are National Payments and SAMOS. As the application of such areas are very specific, it should be determined what voids statistics sourced will fill or value it will add before they are approached. There are some overlaps with regards to questionnaires submitted by financial institutions to other areas in the Bank such as Financial Markets, Financial Surveillance and Banking Supervision. It may be worthwhile to determine the purpose of other surveys elsewhere administered in the Bank to determine the extent it will enrich the asset base of the Research Department. It may therefore be worthwhile to engage with other departments to explore the contributions they can make to enrich or refine artefacts important to the Research Department. This exercise will also enhance any survey consolidation that may be pursued.

5. Assessing uncontrolled information sources

As an intelligence unit, the Research Department not only creates statistics and information, it is also the user of final statistics that are sourced from external agencies and services. Sourcing statistics must be preceded with reason or purpose and if it is fact or direction you wish to enquire about. The usefulness of statistics sourced from external providers is often accepted in good faith. Be circumspect about statistics obtained from paid information services. When subscribing to such services, determine the value, whether it is a convenience proposition from where much statistics can be obtained, or does it supplement or compliment what you already have, or is it a sole provider of statistics. In the absence of control over the compilation of statistics sourced from external parties, it is advisable to employ alternative independent sources that are used to verify the reliability of statistics obtained – a matter of inference. It is also suggested that users are adequately acquainted with methodologies used in the compilation of such statistics. Enquire about the adherence

to standards and best practices by which methodologies are governed. Be alert to unexplained swings in values across periods and breaks in the history of series.

6. Conclusion

Compiling statistics and information is a daunting task and it will continue to be so. With the prospect of new systems with more and new potential to be deployed in the not too distant future, there are many advantages towards '*closing the loopholes*' which are the following:

- 6.1 Most staff members will be involved in the development of new systems, particularly analysis and testing phases. It is an ideal opportunity for everybody to fully acquaint themselves with the offerings of new technologies.
- 6.2 It should enable staff members to work smarter.
- 6.3 It will converge ownership and control activities of data and information assets.
- 6.4 It will greatly enhance the data sourcing capability of the Research Department.
- 6.5 Data mining and business intelligence capability is included in the scope. It should enhance our capability to better assess and analyse data towards '*Statistics that make sense*'.
However, with loopholes being covered, towards '*Statistics that make sense*', there are a few questions that need to be answered:
- 6.6 Are we confident that the basis of our surveys is intact with regards to coverage and representation? Do we know what we don't know or don't we know what we don't know?
- 6.7 Are we confident that the statistics in our control are thoroughly and correctly classified?
- 6.8 Are we confident that in our endeavours to compile statistics, we observe and honour standards and good practices?

Acknowledgements

- 1 Toffler, A., 1970, *Future Shock*, New York: Bantam Books, p 350.
- 2 <http://www.infogineering.net/understanding-information-overload.htm> - 2011-02-07
- 3 Weinberger, D., 2010, Bringing on the info overload, KMWWorld, p 1.
- 4 Wurman, R. S., 1989, *Information Anxiety*, New York: Doubleday, p 222.