

# Data Quality and Compliance with International Standards

*By Edwin St Catherine, Director of Statistics – St Lucia, September 5<sup>th</sup> – 7<sup>th</sup> 2007*

Presented to the **INTERNATIONAL CONFERENCE ON STATISTICS**

‘Statistics and Policy Making in Small Economies: Developing Effective Statistical Systems’

I will look at these issues separately first then I will tie them together in the closing phase of this discussion. A common thread which will run through my comments is the important role which international institutions can play in assisting small countries in addressing issues of Data Quality in a limited way since this is predominantly a matter of institutional effort but more so with being compliant to the most up to date scientific principles, statistical methods and procedures which are internationally accepted. I will use some specific examples to illustrate what I mean by this and encourage further work in all areas of statistics where this is possible.

Data Quality is a pre-occupation of all forward thinking and responsible statistical institutions. In the past, data quality in statistics mainly meant the accuracy of statistical results; we see this orientation in thinking about statistics manifested in our operations even today. This outmoded way of thinking about this issue has damaged the credibility of a lot of statistical work in the past, the release of the 2000/2001 census results is a definite example of this and we are well aware of other similar examples. The modern approach to quality in official statistics refers to a comprehensive, multi-dimensional concept. Apart from accuracy it comprises dimensions of relevance, timeliness and punctuality, accessibility and clarity, comparability and coherence of statistical data releases. This comprehensive quality concept comprises the quality of the whole institution of official statistics; besides the quality of statistical data is the quality of processes, efficiency, the qualification and motivation of members of staff based on a meritocratic approach to managing the statistical agency, the quality of services provided to users and the confidence built and maintained by the statistical agency with the users and other data producers in general.

From a practical perspective a decision that one normally faces is when to release trade statistics by HS 2007 product classification. Lets assume that the trade is normally released within a month for the previous year 2006. If HS 2007 was implemented towards the end of 2006 by the Customs and trade was previously classified in HS 1996, I ask the question, what decisions need to be made to ensure compliance with the release dates given the obvious problems with would have arisen in the data. Indeed every effort must be made to plan for this change so as to effect a smooth transition to the new classification system. However, as occurs too often a decision sometimes has to be made towards the end of January on whether these data can be release given the obvious errors it contains of having two classification systems within the HS product statistics in one year. The question is should it be released? There is to

my mind an assessment of the quality of the information which needs to be made? Of the 5000 or more products identified in the HS 2007 how many are affected by this, how quickly can this be fixed in time for release at the end of January, what is the value of trade affected by this problem in relation to total trade. If it is less than 1% and these errors cannot be addressed before the release date should it be released on the release date with caveats about this quality problem? Unfortunately, too often Statistical Offices would make a decision to delay this release without fully assessing the impact of this problem overly focusing on the accuracy of every detail without realizing that data quality is also largely an issue of timeliness and punctuality.

Another one of our sins against quality statistics is the fundamental issue of building a meritocracy within the Statistical Organization. Often I hear this complaint by managers of Statistical Offices and the argument usually goes like this, because of high staff turnover we are powerless to make substantive and lasting improvements to the statistical system. I will hasten to add this is actually an especially difficult problem the smaller the statistical office. I do not want to downplay its significance from the perspective of situations which are really total out of the control of the manager of the statistical office. The main problem I have with constantly blaming “staff turnover” for the many ills suffered by statistical offices is that it appears to me that the main culprit in creating this problem is how we manage the statistical agency. Very often in speaking to young persons; actually I seek them out, especially those who I know can make a significant impact on the quality of statistics produced; usually their complaints are that this place is not dynamic, I will get stuck here, this place is a dead end, my efforts are not rewarded or respected etc.. My contention is that management of Statistical organizations must ensure that the institution is attractive, progressive, rewards the efforts of the more able, has a calculated risk-taking focus to implementing new methods and continuously invest in training and exposure of staff to the best methods and tools that are available. The more a Statistical agency or any agency for that matter complies with these precepts the less staff turnover it will have and the more effective it will be as an organization. Statistical offices must be helped, encouraged, assisted and if necessary forced to make these changes.

The 2000 Austrian Federal Statistics Act, which entered into force on 1 January 2000, has set a new and modern framework in which official statistics has to be elaborated and disseminated. This new legal framework comprises the quality control of the statistical products, a user orientation, a reduction of the respondents’ burden, efficiency and staff orientation. These principles have been derived from the EU Regulation on Community Statistics (EU, 1997). Apart from objectivity and impartiality, Statistics their Act asserts must adhere to the following principles:

- to apply statistical methods and procedures according to internationally acknowledged scientific principles and standards;

- to control statistics permanently with regard to potential quality improvements;
- to safeguard the timeliness of statistics;
- to minimize the response burden and to provide background information on the survey to respondents and other concerned persons;
- to adhere to specific principles of publication;
- to maintain the confidentiality of individual data.

The issue of “applying statistical methods and procedures according to internationally acknowledged scientific principles and standards” is in itself an aspect of data quality. For Statistical Offices with limited resources, human, financial and technical with limitations in the ability to specialize or conduct significant research and development activities the approach to the implementation of international standards must keep these issues in focus. There are also advantages that our small size creates such as flexibility in implementing change in statistical systems. Our research and development efforts must therefore be focused on keeping abreast with developments particularly as it relates to information technology applied to the statistical field which we can size to suit our needs. International agencies and larger statistical offices are constantly engaged in efforts to integrate improved technology in statistical systems that incorporate updated, internationally acknowledged scientific principles and standards. This is where our opportunities to leap forward lies. In many cases these developments specifically recognize the limitations which small statistical offices have. Let me point to some specific examples of this. Firstly, the IMF’s consumer price index processor specifically recognizes the weaknesses and limitations in current methodological practices with regard to the compilation of Consumer Price indices, particularly as it relates to the use of volatile arithmetic Laspeyres and allows for the implementation of Geometric Laspeyres in combination with three basic compilation methods based on product, geographical area and outlet. Further, the CPI System follows the guidelines of the *Consumer Price Index Manual: Theory and Practice (2004)*, International Labor Office, et al. URL: <http://www.ilo.org/public/english/bureau/stat/guides/cpi/> and incorporates the most recent and updated international standards for the CPI incorporated in the “Resolution Concerning Consumer Price Indices, adopted by the Fourteenth International Conference of Labor Statisticians”. Even if the system is not perfect it is better to work to improve it rather than to re-invent the wheel and not grasp the hands that are out to support us.

Another specific example of this is Eurotrace and we must continue to embrace developments which are occurring with that software as it relates to the processing of trade statistics. The implementation of the Standard for Data and Meta Data Exchange or SDMX has direct bearing on this topic of compliance with international standards. SDMX is gaining increasing use as an effective model for improving the data and metadata exchange amongst data producers, users and intermediary organizations. It was endorsed by the 37th UN Statistical Commission in March of 2006 and further work on its development is now fully embraced by all of the most significant national and international agencies including the IMF, OECD and the World Bank. The question is how do we implement this standard as efficiently as possible? Here again we can find by the use of software developed and supported by the UN, specifically through

DevInfo V5.0 there is a means of integrating international standards and best practice into small statistical offices, in St Lucia our specific application of this technology is dubbed Helen Info 1.0.

DevInfo is designed to facilitate the introduction of internationally accepted and agreed standards for Statistical Data and Meta-data eXchange, without having to deal with the complexity of the issues inherent in the standard itself;

“DevInfo, version 5.0, includes significant advancements in handling metadata. It conforms to standardized information sharing models for metadata storage and dissemination. These metadata standards help achieve efficiency by facilitating data exchange and harmonizing international and national data sets. They decentralize data maintenance, lower the reporting burden on data producers and reduce duplication of effort. They foster use of more timely and better quality data. They shift the process of data distribution from "pushing" data to data users to providing options for data users to "pull" data from data producers as and when new data are published. They allow for data exchange among applications on a broad spectrum of platform.”

while SDMX is based on a broader IT standard for data exchange called XML (eXtensible Markup Language), it is according to ISO based on three statistical meta data standards, namely, standards for data formats, standards for metadata and a registry-based architecture to implement these standards and to exchange data between systems. The other three ISO standards implemented indirectly and directly in Dev Info 5.0 respectively are DDI an acronym for the Data dissemination Initiative for datasets in the social and behavioral sciences, such as household surveys but more importantly the Dublin Core Metadata Initiative, DCMI (ISO 15836:2003) is for a broader range of statistical data resources, such as publications. DevInfo 5.0 stores source information in DDI/DCMI format. The DevInfo Database Administration module is integrated with the International Household Survey Network. The third standard is built into most of the new Geographic Information System software. At the statistical office we have finished a St Lucia settlement map using ESRI's ArcEditor 9 which is the first of its kind, it documents spatially what were previously notionally understood community boundaries on the island. DevInfo 5.0 stores digital map metadata using a standard through which it is possible to communicate with other institutions an implementation of Dev Info linking the maps put into the system with data which can be shared with data providers and vice versa. We have already been sharing much of this resource that we have built with the physical planning unit, the electoral office and more recently with the police to assist in the identification of “crime hotspots”. Dev Info 5.0 will allow us to disseminate spatial data more effectively to other institutions and conversely receive similar input from data producers with whom we collaborate. The isolated efforts of a small statistical office will never be capable of replicating these inputs at the international level or by our bigger siblings and if attempted will require an unacceptably high percentage of our financial resources to adequately address the issues, but I ask to what end? So the idea is to take advantage of this work, contribute to improving it, encourage its growth and assist in its development.

There is however a condition I want to place on this issue of piggy-backing on innovations which are taking place. We must ensure that it is appropriate to our organizations, it must promote the learning and involvement of the staff of our offices, it must engage them and challenge them. Technologies that are so hardwired that we only need to provide input and we get output from a “black box” that we have not had significant participation in creating should be avoided since its adaptability to our circumstances will be in serious question. The process of implementing SNA 93, of developing the supply and use tables for the economy of St Lucia was to my mind a good marrying of international external expert assistance with technology built in excel spreadsheets and significant local knowledge and expertise of staff of my office. Co-opting, expert international, regional or sub-regional assistance is an important aspect of improving data quality and the scientific principles and statistical standards by which we operate.

While I often emphasize the use of IT and ensuring that we keep abreast with technological change in a systemic way within our organizations this is one aspect of it. My emphasis on innovation has as much to do with the packaging of new and improve international standards as it is with issues of data quality from a holistic perspective; however it has also to do with the creation of dynamic and progressive Statistical Offices that our staff can be proud of. It is challenging but I think it is eminently doable.

## **REFERENCES**

Austrian Federal Statistics Act 2000: Federal Law Gazette I No.163/1999, and (amended) Federal Law Gazette I No.136/2001.

COM (2005) 217 final: Communication from the Commission to the European Parliament and to the Council on the independence, integrity and accountability of the national and Community statistical authorities, Brussels.

ECOFIN (2006): 2766th Council Meeting, Economic and Financial Affairs, Brussels.

EU (1997): EU Regulation on Community statistics. No. 322/97, OJ No. L52/1.

Hackl, P. (2004): The Quality Assurance Committee at Statistics Austria. Paper presented at the European Conference on Quality and Methodology in Official Statistics, Mainz, Germany, 24-26 May 2004.

Holzer, W. (2005): Strategic Management in Statistics Austria - Total Quality Management. Paper presented at the High Level TACIS Seminar “Managing Modern National Statistical Systems in Democratic Societies”, Vienna.

[http://www.devinform.org/content/news/news\\_001\\_02.htm?IDX=2](http://www.devinform.org/content/news/news_001_02.htm?IDX=2)