

CHANGING SYSTEM OF STATISTICS EDUCATION FOR OFFICIAL STATISTICS

Ashish Kumar

United Nations Statistical Institute for Asia and the Pacific

kumar83@un.org

Statistics has been playing important role in all fields. Governments have to know the status and changes in social, economic and environmental conditions in their country. Agenda 2030 on Sustainable Development Goals has put additional challenge on Statistical Systems to produce large number of monitoring Indicators. Use of official statistics has been facilitated with faster development in IT. Statistics education has also undergone changes. The training in statistics has become more application based. Statistical Institute for Asia and the Pacific has been providing training in latest statistical standards, adoption of technology in statistical environment and development of analytical skill. Method of delivery of training is also undergoing change with shift towards e-learning. Formal statistics education also must change to address requirement of national statistical systems.

INTRODUCTION

Application of statistics for decision making and policy formulation both by the Government and private corporates have been increasing over years. This increase has been significant in last twenty years with increasing availability of computing power, advent of fast data exchange platforms, expansion of internet and related technologies. This process got further boost because of increased capacity to hold and access large databases, and also storage capacity becoming available at a reasonable and affordable prices. These databases are not only being generated by administrative processes of the Government but also by various operations carried out by private operators specially service providers. Social media has become another important source of such data. To derive knowledge from such large databases, both structured and un-structured, has become an important challenge which is addressed by both knowledge of Statistical methods and ability to analyze the data i.e. knowledge in terms of computer based data processing and development of suitable analytical algorithms.

Purpose of this paper is to understand nature of human resource needs of the National Statistical Systems (NSS) and identify changes required in statistical education system. These changes will be required for new human resources recruited by the National Statistical Offices. There is also need to understand how further training should be given to personnel who are in the system and who have to meet the challenges placed before the system to meet the requirement of monitoring of Sustainable Developmental Goals (SDGs) adopted by UN General Assembly in Agenda 2030 (United Nations, 2015).

OFFICIAL STATISTICS SYSTEMS

The official statistics systems in the Government have been created to meet regular data requirement for its efficient delivery of statistical products and services for planning and policy formulation. The official statistical system has to provide the information on the state of economy, society and environment on regular and periodic basis so that policies implemented may be evaluated and new policies developed to meet new challenges. Apart from some administrative data generated through administrative system, official statistical systems have been traditionally based on the data collection through surveys- both Censuses and Sample surveys. Sample surveys based on probabilistic sample selection ensure properties of independence and representativeness of the sample from the population and the estimates of the population parameters are mostly unbiased with least variance ensuring high degree of precision. Later, in some National statistical offices there has been shift for using information based on administrative registers maintained by the Governments for some of their administrative or legal purposes. There has also been an effort to generate administrative/ business registers in national statistical offices based on the information available from tax administration systems in the country. Many a times these registers are used for drawing of sample and hence obtaining estimates using the information in the register as frame (Statistics Sweden, 2017).

Challenges in generation of official statistics

Official statistics generated through sample Survey based information came under a lot of challenge due to four main reasons:

- (i) Sample survey theory assumes that it is always possible to have a complete frame of all units without any duplication or omission. In reality this hardly happens.
- (ii) Scientific sample survey always assumed that all the samples will be surveyed which in real life does not happen. There has been significant increase in lack of response in sample surveys over years. Theories developed to address the lack of response have been mostly inadequate to handle the issue. This has been adversely affecting the quality of estimates being produced with the help of surveys.
- (iii) The cost of collection of information from a scientifically designed sample surveys has been enormous. This cost has been increasing with increasing cost of the manpower which is usually deployed to canvass questionnaires.
- (iv) Time taken in collection of information and processing of data has been a serious issue. The information from such surveys becomes available after considerable time lag. This has been addressed to some extent by use of Computer Assisted Survey Data processing techniques.

Above mentioned conditions have led to a situation where there is now increasing emphasis to look for data from other sources. This means increasing use of the data available from the mechanisms which are not necessarily for the purpose the data is likely to be used (Statistics Finland and Eurostat, 2017).

Changing role of national statistical systems (NSS)

With the increasing computerization and expansion of internet, there have been regular and quick demand from policy makers and users of data for the information or indicators which are relevant for policy formulation in the Government and also on issues relating to public debate. This expectation has risen to the extent that the policy makers and data users demand the statistics and indicators without any delay. Further, these demands are putting additional pressure on the national statistical systems to modernize and look for new data sources to meet the demand in timely manner.

Adoption of Agenda 2030: Transforming our World by the member states of the UN has put additional responsibility and challenge to the national statistical systems to produce the indicators to monitor progress of 17 Sustainable Development Goals, 169 targets for these Goals and 232 global Indicators (further disaggregated by various characteristics and class of populations) identified so far. Producing all these indicators has been a challenge not only for developing countries but also for developed countries. Countries like Japan have been able to produce hardly 40 percent of all the indicators which includes a number of proxy indicators.

Another issue faced by the statistics community in terms of Agenda 2030 is that the implementation of SDGs will be with an objective to 'Leave no one behind'. In statistical terms, this means disaggregation of all the indicators by Income, Sex, Age group, Geographical location (Rural / Urban), Race, Ethnicity, Migratory status, Disability etc. It has to be noted here that present official statistical systems are not geared to produce statistics and indicators with such a level of disaggregation. The Chief Executive Board of the UN agencies have endorsed shared framework for action on combating inequalities and discrimination to United Nations system support for implementation of 2030 Agenda. [See United Nations Chief Executive Boards Coordination (2017)].

The National Statistical System is under tremendous stress to meet both the challenges i.e. producing large number of indicators and also producing these indicators with various levels of disaggregation. It is not possible for the system to start producing all the information just by using traditional methods, censuses and sample surveys, of data collection, compilation and analysis. The perception and approach for collection, processing and analysis of data and statistics in national statistical systems has to undergo a major change. This is often referred as 'Data Revolution'.

Systemic changes in NSS

Role of NSOs has to undergo a major change as a result of new demands especially in the context of Agenda 2030. NSOs, which used to function in isolation based on a few statistical business processes following internationally defined standards, have to become coordinators of all the statistical information to be produced by the whole of NSS. The following systemic changes are required in NSS to meet the challenge.

The first step in this regard is to understand and identify new data sources within the Government which have not been tapped so far. NSO/ NSS have to be given authority by the Government to access all such data. The Governments have to remove various bottlenecks which exist in the context of privacy of information in sharing of data across the departments within Government. This is a major challenge since this will also affect the complete ownership of the data of some departments. These departments may have apprehensions that sharing of such data may result into scrutiny on their actions by other departments in the Government. Such doubts may have to be dispelled by the Governments and NSOs.

Secondly, the NSOs / NSS at present do not have authority or wherewithal to access the data which are in the custody of private sector. There is a need to develop laws/ regulations to make such data accessible to NSO/ NSS so that they develop familiarity of the data, understand its usefulness and start using it to the extent it is possible to use. This also will give an opportunity to NSS to triangulate this data with the conventional data available with them from the existing traditional sources.

Thirdly, it is also important for statistical offices to work with the experts with different domain knowledge. This requires development of system by which the statistical personnel work in collaboration with other discipline experts. This requires change in the attitude of the statistical personnel and also making them ready to understand the language of experts from other disciplines.

CHANGING SKILL REQUIREMENT OF STATISTICAL PERSONNEL IN NSS

Changes which are taking place in the NSSs and NSOs require development of different set of skills of the personnel in these organizations (Jeff Wu, C.F., 1998). The demand of the system will have to be met by bringing in new set of persons who are often referred as data scientist and also retraining of the existing conventional statisticians to the new methods. Some of the specific suggestions in this regard are:

- (i) Bringing in professionals with data science background in the National Statistical systems
- (ii) Training of young statisticians in the data science related fields.
- (iii) Development of systems so that professionals outside of the Government and also personnel of other Departments can work with the data available in NSS to analyze and use it for meaningful purposes for policy and also monitoring of various policies/ programmes of the Government.
- (iv) The system has to become innovative in use of new data sources and new methodologies. There has to be encouragement to do so with in-built system to protect personnel in case such data is not found useful and productive.
- (v) There is a need to develop communication and advocacy skills in the national statistical systems including statisticians in these systems, so that it is possible for them to effectively communicate the results of their analysis to persons with no or almost no knowledge of statistics. It is also important for generating investment in statistics from the rest of the Government specially policy makers.

Many national statistics office have recognized the need (German Federal Statistical Office, 2017).

STATISTICAL EDUCATION AND CHANGES INTRODUCED

Conventional courses in Statistics in the universities primarily focused on mathematical statistics with some application of these methods in real life situations. These conventional courses have been producing statisticians with strong theoretical background with less of application knowledge. These courses started undergoing changes after easy availability of computers. Many

of these courses introduced computer application as a part of their syllabus. However, this is not sufficient to meet the present day challenge on national statistical systems.

Many universities and academic institutions have introduced new courses which are called “Business Analytics” or “Data Science” courses. These courses have come in high demand due to their ability to provide basic skills to students on statistical methods and also in analytical skills using computer based databases and use of various software. Two more important subjects have been included in these courses. One of the most important feature, which is added, in the business analytics course is the transfer of knowledge of business practices which helps students to ask relevant context specific questions and search for their solutions based on their knowledge of statistics and data analysis. Second important aspect in both these courses, which has become important in the present context, is the ability of analyst to present results in such a manner that are understandable to the client. This requires development of communication skills of such analysts or data scientists (Massimiliano Marcellino, 2017).

Further changes required to meet the challenge faced by the NSO/NSS are development of courses where apart from the knowledge of statistics and data analysis, there is also components relating to the organization of the Government, data getting generated within the Government and also the policy questions which are dealt in the Government. This component of the course may become more specific to a country or a group of countries so that data analyst or statistician is in a position to appreciate the issues involved, raise questions in the correct perspective and try to find solution from the data available in the Government or from other sources and look for solution to address the problem.

Apart from introduction of such courses in the universities and other academic institutions, there is a need to develop training facilities in the Government to retrain the existing staff by providing them skill which are needed in the present context.

It is not only the analytical skills which are important for meeting the challenge, there is need for development of managerial skills both of the systems and also databases for addressing the challenge from the new environment and therefore to fully transform the National Statistical Systems to a modern and fully computerized system [see, for suggestions, Central Bureau of Statistics of Israel (2017)].

Role of SIAP in the context of UN agenda 2030

Statistical Institute for Asia and the Pacific (SIAP) has been set up as a subsidiary body of UN Economic and Social Commission for Asia and the Pacific. The institute has a mandate to strengthen capability of member countries/ associate members and economies in transition to collect, analyze and disseminate statistics and to produce timely high quality statistics for economic and social development planning. It also has to strengthen the statistical training and related activities in the member states of ESCAP. In the context of monitoring of Sustainable Development Goals (SDG) resolved by UN General Assembly and global indicators identified by the UN Statistical Commission, the role of SIAP is to strengthen the capability of the member states/ associate member states to produce high quality statistics on the indicators for monitoring of progress of SDGs. Role of SIAP is catalytic in nature in terms of building capacities of the statistical offices in the region. It provides training to trainers of the national statistical offices or the officers of the concerned Ministries in the areas which are new and are challenging for many of the national statistical offices to handle at their own [see, for full description, United Nations Statistical Institute for Asia and the Pacific (2014)]. For achieving these objectives, the SIAP has been undertaking following types of training programmes:

- (i) Trainer’s training
- (ii) Face-to face training (both long term and short term training programmes)
- (iii) E-learning programmes
- (iv) Blended learning (Most of the training programmes are becoming blended learning programmes)

The institute has been reorienting its training programmes by suitably bringing in the link between statistical business processes with the global and national indicator frameworks for Sustainable Development Goals of Agenda 2030.

There is a need to transform the work of national statistical offices to meet the requirement of the SDG indicators by modernizing the statistical business processes. This requires adoption of ICT tools at all stages of statistical processes. Practical training on use of Computer Assisted Personnel Interviewing (CAPI) software for electronic data capture has been a major step to ensure availability of quality and timely survey data for the purpose of SDG monitoring. The institute has been providing training on various software tools for electronic capture of survey data.

The Institute has also started a training course on data disaggregation to ensure that the countries develop capability to address the issue of disaggregation in their context for reporting on SDG related indicators. The course apart from data disaggregation of sample survey data also focuses on use of small area estimates for the purpose.

The institute shall be organizing courses related to use of big data for compilation of SDG related indicators specially building on the experience from Republic of Korea and Japan so that other countries in the region may adopt these practices in due course. These courses will use Satellite imagery data, mobile data, credit card data and the data available from scanners.

There is increasing emphasis on development of leadership and management skill of the heads and chief statisticians of national statistical offices. The institute apart from organizing leadership training with various partnership agencies also organizes Management Seminar once in two years for improving leadership skills of the senior members of National Statistical offices including heads of NSOs. Further, there is also need to develop National Statistical Development strategy (NSDS) integrated with the Development Plans of the countries for compilation of SDG Indicators. Such courses are being organized in collaboration with PARIS21 and would be continued to be organized in future.

The institute has developed a course on Climate change statistics which apart from other indicators provides training on compilation of System of Environment Economic Accounting (SEEA) and Frame work for Development of environment Statistics. This will be further expanded to cover Ocean related statistics and Indicators. Gender empowerment has been a major theme of all countries including of UN. There has been considerable effort to empower women. This empowerment is possible only if the data, statistics and indicators in global SDG monitoring are disaggregated by sex. Quality of statistical indicators is a critical issue for properly tracking the progress of SDG related indicators and also ensuring inter-country comparisons. The institute will organize a course to address quality issues related to data used for compilation of SDG related indicators. The institute has been laying a lot of emphasis on developing e-learning courses on various issues like Introduction to Official Statistics to the beginners in the context of SDG Indicator frame work, SDGs, targets and related monitoring Indicators, System of Environment and Economic Accounting, System of National Accounts, Compilation of price statistics including price related SDG indicators, Use of administrative data for preparation of Statistical Business registers etc.

CONCLUSION

It is true that National Statistical Systems are under tremendous pressure to meet the challenge of producing large number of Global and National Indicators for tracking the progress of SDGs and other related indicators. A fast transformative action is needed to strengthen and empower these systems by the Governments to ensure that they meet the challenge. Statistical systems will also have to take this an opportunity to modernize its processes and start using new sources of data to meet the requirement. Apart from systemic changes in national statistical systems, there is a big challenge to get personnel with required skills in the national statistical systems. This has to be done both by recruiting personnel with required skills not only in statistics but also in data science. There is also need to build capacity of the personnel in the system.

There is a need to introduce the following two components in graduate and post graduate courses in Statistics:

- (i) Introduce a course on official statistics, which has become an important discipline on application of statistical methods. This would also require introduction on functioning of the Governments, their data requirements and sources from where the data generated is used for various purposes.

- (ii) Introduction of data science analytical capabilities in the programmes including their suitability in statistical theoretical background.

The second part is also relevant for building of capacity of presently available manpower in statistical systems. NSSs will have to take lead in training of their staff in analysis of big data and management of databases.

SIAP has been playing a catalytic role in bringing about the change by training a few selected personnel in the NSOs/NSSs in the region. These trained personnel in SIAP will have to take the role of trainers to further train the personnel in NSOs/NSSs in their countries. But to effectively meet the challenge, the NSOs will have to take lead in terms of developing capabilities of their personnel and creation of an environment so that personnel take lead in using new sets of data to meet the challenge of SDG indicator production.

REFERENCES

- Central Bureau of Statistics of Israel (June 2017). *Statisticians or data scientists? Future of official statistics in the era of new technologies and new data sources*. Note for the *Sixty-fifth plenary session of the Conference of European Statisticians*, Geneva. Retrieved from <https://www.unece.org/index.php?id=43851>
- German Federal Statistical Office (2017, June). *Statistics education in times of big data*, Note for the *Sixty-fifth plenary session of the Conference of European Statisticians*, Geneva.
- Massimiliano Marcellino (2017, June). *Big data skills: Challenges for University World: Creating a new generation of data scientists*. CES 2017 Seminar on the new generation of statisticians and data scientists, Geneva. Retrieved from https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/2017/PPT/20June_09h45_4a_Big_Data_skills_Marcellino.pdf
- Statistics Finland and Eurostat (2017). *Skills for new generation of statisticians*. Paper presented in Seminar on the next generation of statisticians and data scientists for the Sixty fifth plenary session of the Conference of European Statisticians, Geneva. Retrieved from <https://www.unece.org/index.php?id=43851>
- Statistics Sweden (2017). *Statistics production system 4.0*. Note presented at the Sixty fifth plenary session of the Conference of European Statisticians, Geneva. Retrieved from <https://www.unece.org/index.php?id=43851>
- United Nations (2015). *Transforming our world: the 2030 Agenda for Sustainable Development*. in the resolution adopted by the UN General Assembly on 25 September 2015, New York. Retrieved from http://www.un.org/en/development/desa/population/migration/generalassembly/docs/globalcompact/A_RES_70_1_E.pdf
- United Nations Statistical Institute for Asia and the Pacific (2014). *Who We Are*. Retrieved from <http://unsiap.or.jp/about/index.html>
- United Nations System Chief Executive Boards Coordination (2017). *Leaving No One Behind: Equality and Non-discrimination at the Heart of Sustainable Development: A Shared United Nations System Framework for Action*. Retrieved from <https://www.unsceb.org/CEBPublicFiles/CEB%20equality%20framework-A4-web-rev3.pdf>
- Wu, C. F. J. (1998). *Statistics=Data Science?*. P.C. Mahalanobis Memorial Lectures, 7th series. P.C. Mahalanobis Memorial Lectures, Indian Statistical Institute. Archived from the original on 29 October 2013. Retrieved 18 August 2013.