Integrated Statistics: A journey worthwhile

The 2030 Agenda for Sustainable Development integrates three dimensions of development (economic, social and environmental) into an inclusive framework aiming to transform the world into a better place for all, for today and the future. Monitoring, planning and analysis across these dimensions requires a vast range of quality, timely and reliable data and statistics which are currently beyond the reach of National Statistical Systems (NSSs). To meet increasing and evolving demands, NSSs need to transform. The Asia-Pacific statistical community, in its collective vision and framework for action1, has introduced five action areas for transformation, one of which is integrated statistics for integrated analysis. This paper discusses integrated statistics, how integrated statistics can help transform NSS, and how it can be delivered.

What are integrated statistics?

Integrated statistics can be defined as the ultimate outcome of an integrated system of producing and using statistics.

Integration may take place at different dimensions. In December 2017, a group of national and international experts in Asia-Pacific agreed upon four different dimensions of integration: a) process integration, b) data integration, c) conceptual integration, and d) disciplinary integration[2].

Process Integration	Seeks to reduce process duplication through centralization and consolidation of common statistical processes.
Data Integration	Aims to explore ways to reduce data duplication and increase data usability, findability, and comprehensiveness.
Conceptual Integration	Focuses on aligning multidisciplinary concepts and involves in restructure of existing data to common concepts, classifications and methods.
Disciplinary Integration	Explores integration that requires convergences among disciplinary concepts and make a common language.

This paper aims to shed more light on process integration and data integration.

How can integrated statistics help?

To respond to the increasing demand for more disaggregated and timely data in the face of limited resources and declining cooperation from respondents, statistical systems need to seek an efficient way of production without compromising on quality. Integration can help by minimizing duplication, reducing respondent burden and increasing cost-efficiency. Integration can also contribute to improving various quality dimensions of official statistics.

---
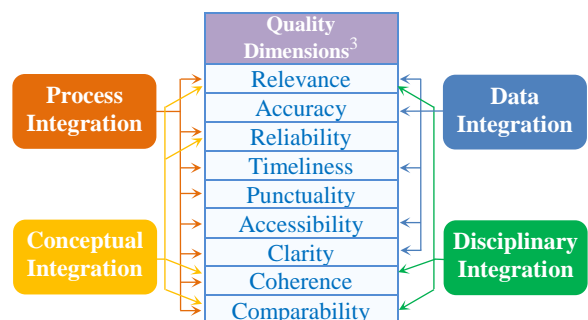
[1] Advancing official statistics for the 2030 Agenda for Sustainable Development: a collective vision and framework for action by the Asia-Pacific Statistical Community (https://www.unescap.org/sites/default/files/pre-ods/E.ESCAP_.CST(5).1.Rev_.1.Collective_Vision.English.pdf)

[2] https://www.unescap.org/sites/default/files/Dimensions_of_integration_in_the_statistical_system.pdf

[3] Quality dimensions are based on the draft version of United Nations National Quality Assurance Framework Manual for Official Statistics (as of May 2019), https://unstats.un.org/unsd/dnss/docs-nqaf/UN_NQAF_Manual-Unedited_manuscript_of_3_May_2019.pdf

Moreover, integrated statistics can add value to open data[4] Combining open data can reveal new insights and convey new valuable messages relevant to informed decision-making. "Open Data Principles"[5], especially principles which mandate data to be timely, comprehensive, accessible, usable, comparable and interoperable support integration.
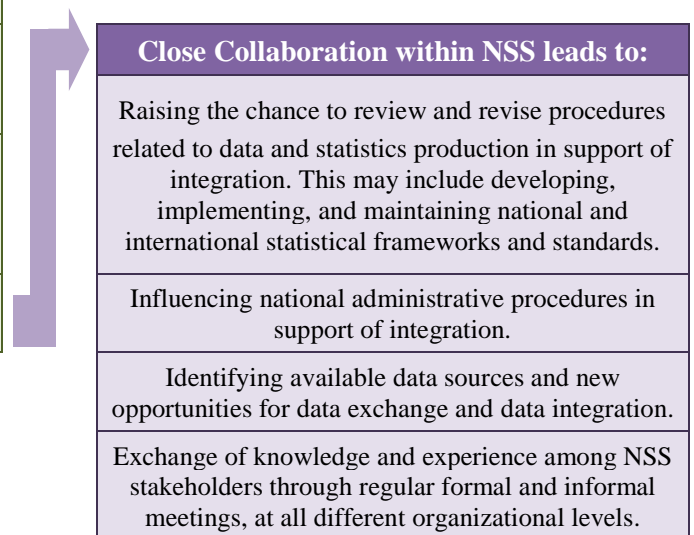
What are main enablers?

Transforming statistics production from independent production lines into an integrated system requires both technical and organizational improvements. Two key requirements to put such a system in place are transformed governance environment and capacity development.

Transformed governance environment

Transforming a governance and institutional environment may require revising National Statistics Act, developing legislative frameworks, and establishing formal and informal

coordination mechanisms. These should lead to a) strong and sustainable support from high-level policy makers, b) empowered leadership role of National Statistical Office (NSO), and c) partnerships and intensive collaboration among all stakeholders within and outside the government.

A transformed governance and institutional environment accelerates NSS move towards integrated statistics.

Capacity development

It is the people who transform systems. Staff working in statistical units of governments need to develop new skillsets that enable them to harness new technologies, apply complicated statistical methodologies, understand complex policy issues, and effectively communicate and negotiate. Beside comprehensive capacity development programs, part of technical capacity building can be fulfilled through regular and structured user-producer dialogues.

In Transformed Governance Environment:	
	NSS activities are prioritized based on national development programs and are fully funded.
	NSO is the coordinator of statistical system. Roles and responsibilities of all other actors and stakeholders are also clearly defined.
	NSO is mandated to collect data from different sources for statistical purposes; thereby, has access to data and statistics produced by other national agencies and private sector.
	Close collaboration of NSO and other members of statistical system is in place.

Close Collaboration within NSS leads to:
Raising the chance to review and revise procedures related to data and statistics production in support of integration. This may include developing, implementing, and maintaining national and international statistical frameworks and standards.
Influencing national administrative procedures in support of integration.
Identifying available data sources and new opportunities for data exchange and data integration.
Exchange of knowledge and experience among NSS stakeholders through regular formal and informal meetings, at all different organizational levels.

How to deliver?

As previously mentioned, this paper focuses on process integration and data integration and addresses delivery issues relevant to each.

Process integration

Process integration seeks to reduce process duplication through centralization and involves consolidation of common statistical processes. Statistical tools that support

process integration include Generic Statistical Business Process Model (GSBPM), the Common Statistical Production Architecture (CSPA), the Statistical Data and Metadata Exchange (SDMX), the Generic Statistical Information Model (GISM), and the Data Documentation Initiative (DDI).

Process integration encourages developing and applying common tools and procedures. It inspires standardizing and harmonizing all factors that incorporate the production of

---

[4] "Open data is digital data that is made available with the technical and legal characteristics necessary for it to be freely used, reused, and redistributed by anyone, anytime, anywhere", International Open Data Charter (https://opendatacharter.net/wp-content/uploads/2015/10/opendatacharter-charter_F.pdf)

[5] Open data principles mandate that data should be: 1. Open by Default 2. Timely and Comprehensive 3. Accessible and Usable 4. Comparable and Interoperable 5. For Improved Governance and Citizen Engagement 6. For Inclusive Development and Innovation.

statistics including concepts, definitions, classifications, measurement frameworks, measurement instruments and statistical methodologies.

Official statistics are mainly produced through census, survey, administrative data, and increasingly new sources such as big data. Despite common ground among them, each source deals with different issues regarding process integration.
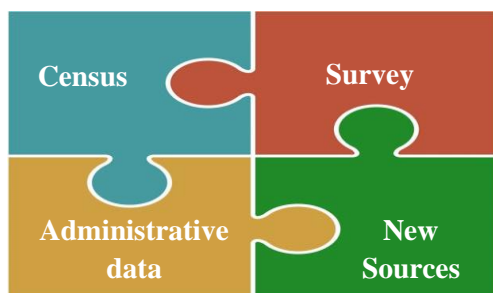
Surveys

In comparison to other sources of data, surveys are more flexible in design and conduct, thus easier to be influenced in support of process integration. Following are some of the mechanisms that facilitate process integration across surveys within an NSS.

NSO as the coordinator of NSS can:
Establish inventories and develop tools and resources that enable all stakeholders within NSS to apply common concepts, definitions, classification and measurement frameworks in surveys.
Develop sampling frames to be used in surveys conducted by stakeholders in NSS. Establishing and using a *common frame*[6] or a *master sample*[7] a) harmonizes survey population and sample design across surveys, b) increases cost-efficiency, c) facilitates organizing field operation, d) enables monitoring of respondent burden, and e) promotes potential for integration of data collected in different surveys.
To encourage use of sampling frames, NSO should provide a procedure through which stakeholders can benefit from sampling frames, without breaching confidentiality.
To the extent possible, replace individual surveys with multi-purpose integrated surveys, such as Integrated Household Surveys which are cross-sectoral surveys, made of multiple modules.

Censuses

The huge operational scale and cost of a census limits its flexibility to be influenced in support of process integration. However, there are still provisions that facilitate process integration.

To facilitate integration, NSO can:
To the extent possible, adhere to common concepts, definitions, classifications, and geographic boundaries, or at least keep the possibility of mapping among different censuses.
Apply standardized national addresses or ideally geo-referenced addresses, to facilitate integration with administrative data as well as new sources of data.
Include national identification number of units such as personal identification number (PIN), and business identification number (BIN) to facilitate integration among censuses as well as integration with surveys and administrative data.

Administrative data

Administrative data, originally collected for non-statistical purpose within an administrative agency, has less potential of being influenced in support of process integration unless there exist enough mechanisms which assure a) coordination and collaboration between NSO and administrative data holders, b) NSO's access to administrative data, and c) compliance with the national and international standards and classifications.

NSO can benefit from encouraging administrative agencies to make slight adjustments to enhance coverage and quality of their data, to improve alignment with standards and classifications, to migrate from paper-based to computerized databases, and to switch from decentralized sub-national systems to centralized national systems. However, as a sustainable solution, NSSs could develop an integrated register system where national administrative databases can be linked to each other and made available for integrated production and use of data and statistics.

---

[6] Common frame, from which business survey frames are derived, comprises list of all statistical units that are active, or potentially active, or were active within the previous reference year. (https://www.unece.org/fileadmin/DAM/stats/publications/2015/ECE_CES_39_WEB.pdf)

[7] Master sample is a sample from which subsamples can be selected to serve the needs of more than one survey or survey round (https://unstats.un.org/unsd/publication/unint/DP_UN_INT_84_014_5E.pdf)

One model for such a register system is illustrated in figure 1, which is a simplified version of the model introduced in Wallgren (2014)[8]. In this model, three base registers are established for three main units: persons, businesses (i.e. enterprises, establishments, etc.), and locations (i.e. buildings, dwellings, etc.). Base registers that aim to be a complete list of units are linked to each other by national identification numbers, including personal identification number (PIN), business identification number (BIN), and location identification number (LIN). This system permits all other administrative sources related to persons, businesses and/or locations to be linked to base registers and enables producing register-based statistics in a coherent way.

Minimum requirements to establish such a model are existence of a) adequate legal framework that assures NSO's access to administrative registers kept by public authorities and protects confidentiality as well; and b) unified systems of identity numbers that are used in all administrative systems in a way that the same identity number follows a unit over its lifetime.

New sources of data

Compared to traditional sources of official statistics, i.e. census, survey, and administrative data, big data can come with lower cost and can lead to more timely and granular statistics. However, there are still issues that hinder exploiting such data, such as potential of bias, uncertainty about accuracy of user-generated contents, difficulty of coordination with data holders, tight privacy regulations and social license.

Building partnerships at national, regional and global levels can be a key to address some of the issues. Data collaboratives, for example, facilitate public-private data sharing for public good and can assure privacy and restrict misuse of data. Part of the technical obstacles can be tackled by capacity development on using these new sources and integrating them with traditional sources.
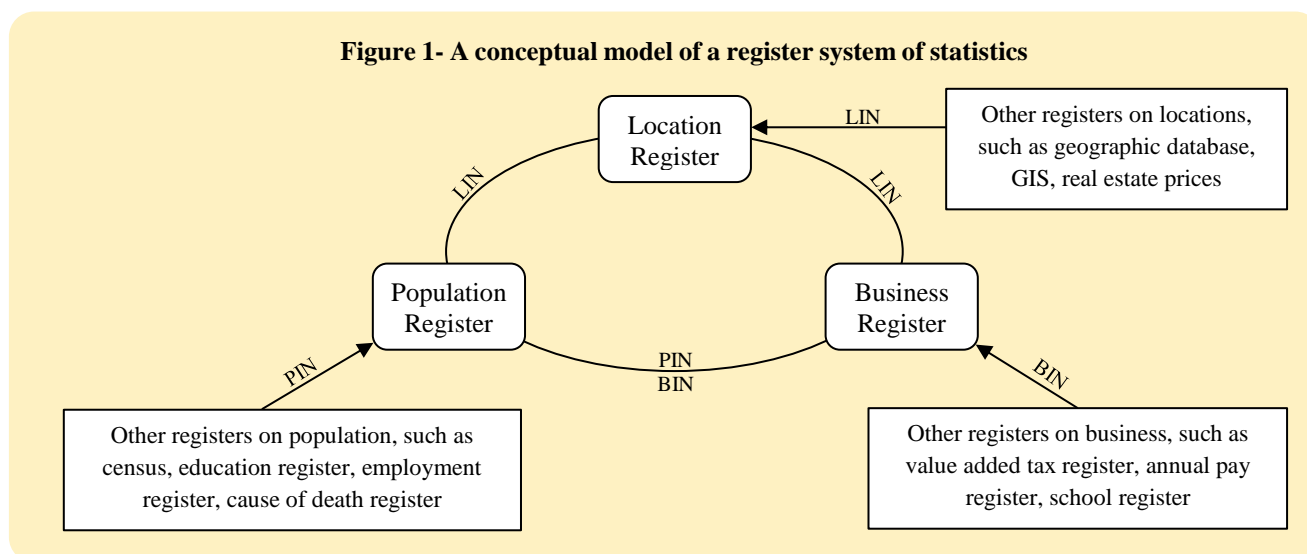
Data integration

Data integration aims to explore ways to reduce data duplication and increase data usability and findability. Statistical tools to be used are metadata, inventorying, archiving, modelling, and the National Spatial Data Infrastructure (NSDI). Process integration can greatly facilitate data integration.

Data integration is defined as linking data from separate data sources, designed and collected primarily with no intention of being used together[9]. Data matching or linking is a primary step for data integration.

Data matching requires:
Legal authority and social license
Access to data and metadata of each source
Interoperability in terms of consistency in concepts, definitions, classifications and data formats
Technical capacity

Figure 1- A conceptual model of a register system of statistics

---

[8] Wallgren, A. & Wallgren, B. 2014. *Register-based statistics: statistical methods for administrative data.* 2nd ed. Chichester: John Wiley & Sons, Ltd.

[9] http://archive.stats.govt.nz/about_us/legisln-policies-protocols/data-integration-gdlns.aspx

Access to data and metadata can be provided within a transformed governance environment, as previously mentioned. Adopting international standards such as SDMX[10] enables transmitting or sharing data and metadata within NSS in an efficient way.

Use of the SDMX leads to well-structured data and metadata in reusable formats and creates interoperable data and metadata.

Besides incompatibility of data formats, in data integration, official statisticians face other issues related to lack of harmonization, such as mismatch of statistical units, reference time, etc. Dealing with such conflicts requires knowledge and experience. Availability of well-structured metadata assists in identifying mismatches and the way to resolve them.

Moreover, data integration involves manipulating different types of data sources. Combining different sources is not often easy due to lack of unique key, and more complicated record linkage methods may be required. Thus, official statisticians need to build skill and develop capacity in data science and record linkage techniques through training courses and practical exercises.

Improving data availability

Once data from different sources is integrated, there are methods that can be applied to improve availability and quality of disaggregated statistics.

Some of the common methods are:

Imputation methods to handle missing values resulted from data matching.

Small Area Estimation (SAE) to provide reliable estimates for variable(s) of interest where the available information is not sufficient for a valid estimate.

Synthetic data generation to artificially create data by using models based on population distributions in order to strengthen the base for more disaggregated statistics.

Enhancing data accessibility

Integrated data can be made accessible to users. In this regard, some countries have established data hubs, where data from different sources are linked, integrated, anonymized and made available to users together with interactive analytical tools which enable users to develop and visualize their own outputs. One example is StatNZ experience that has integrated data from different sources and developed two databases, named Integrated Data Infrastructure (IDI) and Longitudinal Business Database (LBD) which hold anonymized microdata about people and businesses, respectively.

Some countries have established integrated dissemination platforms, where official statistics produced by different national agencies can be accessible in one platform working as one-stop-shop for official statistics. Their difference from national data hubs is that in these platforms, data from different sources are not necessarily linked. One example of such a platform is the Unified Interdepartmental Statistical Information System (UISIS) which is the main supplier of official statistics in the Russian Federation at federal and regional levels. Besides public authorities, UISIS provides other users with free access to official statistics.

At the global level, the Federated Information System for Sustainable Development Goals is a platform creating an enabling environment for national and international reporting, integration and analysis of SDG data and statistics across a system of federated data hubs. This SDG reporting platform is acting as a strong incentive for NSOs to establish their own national dissemination and reporting platforms.

Way forward

This paper reviewed different aspects of integrated statistics, together with opportunities and challenges NSSs face. Integrated statistics is a worthwhile journey for NSSs despite these challenges. Now it worth asking how global and regional statistical community can support NSSs in their journey to integrated statistics for integrated analysis?

---

[10] Statistical Data and Metadata eXchange

Global and regional possible approaches to support countries in adoption of integrated statistics

Establishment of a mechanism to support regional and national efforts in delivering integrated statistics. This can be in the form of establishing communities of practice as decided by ESCAP Committee on Statistics in its sixth session in 2018[11]. The communities can be formed through widely distributed calls for expression of interest, managed electronically, and using modern technology and online platforms. The main objective of communities would be knowledge generation, sharing lessons and good practices, and identification of replicable strategies.

Building partnerships to mobilize resources and technically support national requests for adoption of integrated statistics. These partnerships may lead to developing guidelines, tools or modules assisting integration, or providing technical assistance to countries. One example is MICSLink, an initiative which is being conceptualized by UNICEF. It aims to provide a facility to countries which choose to link data from MICS[12] survey to administrative sources. Such facilities can also be implemented in partnership with other agencies.

Establishment of an online platform that provides a suite of learning opportunities, including e-learnings, on general and specialized issues of integration. It can also provide a sandbox for data integration, small area estimation, synthetic data generation and other related practices in which national statisticians can experiment with data without requiring deep knowledge on data science or complicated statistical methods. The UN Global Pulse and UN Global Platform are two such examples.

For more information regarding ESCAP's work in statistics development please visit:
http://www.unescap.org/our-work/statistics

Previous issues of Stats Brief:
http://www.unescap.org/resource-series/stats-brief

Contact us: stat.unescap@un.org

---

[11] https://www.unescap.org/sites/default/files/ESCAP_CST_2018_6_Report_of_the_Committee_on_Statistics sixth_session.pdf

[12] Multiple Indicator Cluster Survey