The production process in official statistics

The production of official statistics is a complex chain of operations. It starts with the investigations about information needs of various users, their filtering and subsequent bundling in such a way that one activity of official statistics can generate results that fulfil a great number of user needs and are not targeted exclusively to one user group. This phase is normally not carried out every year in a systematic and comprehensive way, but only when a multi-year programme is set up; annual adjustments are possible, but mostly only partial. Information needs have to be investigated in terms of results or outputs; the translation of these needs into the best way of collecting data from respondents, including the arbitrage between the secondary use of administrative or similar data collected outside the statistical system, and specific data collection for the purpose of official statistics though statistical surveys and censuses, is the core task of the statistical system in this programming phase. The programming phase is iterative, since it involves balancing of needs with available resources and priority setting.

Once the statistical objects and the sources for obtaining information for them has been fixed, and the information needs to be covered, allocated to them, each statistical survey, and each secondary use of administrative or similar data, has to be designed/redesigned, tested, and the tools and resources necessary for full implementation prepared and adjusted if necessary following the test. The design phase also includes the definition of the results to be published as official. In the case of statistical surveys, the data collection phase is a crucial part in the design of operations that are under the responsibility of a statistical producer; in the case of administrative or similar data, this phase in outside the statistical system, with statisticians hopefully being consulted in the process of decisions made about their structure and content. For statistical surveys, the data collection phase itself is a key phase in terms of management and use of resources of the NSO.

Once the data have either been collected through statistical surveys, or been handed over to the NSO or another statistical producer in the case of administrative or similar data, the processing phase includes the data entry, control, coding, editing and in some cases imputing of the unit-level data, with the possibility of matching with other sources, and the aggregation of unit-level data to the pre-defined official results (including the necessary quality parameters). This phase is particularly IT-dependent; some of the processing may even be merged with the data collection phase if CATI or similar computer aided techniques, or electronic reporting, are used. The results and the quality parameters have to be analysed carefully before being cleared for the next phase; this may involve integration or at least systematic comparisons with other sources about the same phenomenon at the aggregate level.

The phase of dissemination is more than the release of the pre-defined results in various forms (press releases with comments; hard-copy publications; electronic dissemination on the internet and in other various forms); it may include subsequent publications with more detail or analytical content, or for specific user groups, and it includes the generation of additional results for specific user requests (statistical services). For these purposes,

the final set of unit-level or micro-data have to be stored and well documented for a considerable period.

As a last phase, the whole process has to be evaluated in order to identify and address possible improvements in efficiency and quality, which are then either fed into the next wave or considered at the next systematic programming exercise.

In addition to the production processes, a statistical system needs a number of support processes of cross-cutting character. They address resources and the statistical infrastructure. No statistical system can work without well motivated and professional staff, and without a carefully designed IT infrastructure targeted to the tasks of official statistics (which are different in focus from the type of transaction-oriented tasks normally met in the rest of the public and in the private sector, notably with respect to the complexity of metadata).

Examples for processes that address elements of the statistical infrastructure, and do not completely follow the pattern outlined above for production processes, are statistical registers and international cooperation

Since users cannot replicate easily this complex chain of operations themselves, they have to trust the results that are published as authoritative and unbiased. But trust in the results has its root in the confidence that producers are professionals, and that the institutional framework in which they operate allows them to act professionally in all situations, even when results of official statistics are bad news for some actors in the political scene. This is the root of principles like professional independence and impartiality. Decisions on the choice of sources, the methods to be used for collecting data and compiling results for official statistics, the selection of results to be disseminated as official, and all decisions about timing and forms of dissemination, have to be strictly internal to the statistical system so as to be free from any interference that could bias such decisions in order to distort or hide results in a certain way. As a corollary to professional independence, all the methods used have to be fully transparent. Relying on recognized international standards and good practices in national statistics across the globe is an important way to ensure trust and be recognized as professional.