

Management of Quality in a changing data ecosystem: the case of FAO

Ngarsaim Espoir Beram¹, Valérie Bizier², Clara Aida Khalil³

¹Associate Professional Officer, Food and Agriculture Organization of the United Nations, Italy

²Senior Statistician, Food and Agriculture Organization of the United Nations, Italy.

³Statistician, Food and Agriculture Organization of the United Nations, Italy.

Abstract

The data ecosystem is being increasingly affected by the emergence of new data sources, reshaping statistical business processes and data and statistics quality management. Indeed, these dynamics involve opportunities for statistical advancement (increased timeliness, granularity, strengthened decision-making) as well as concomitant risks (data privacy, data integrity and statistics quality, data access and sustainability of data sources) that need to be addressed.

To ensure that the FAO remains relevant in this rapidly changing environment, a paradigm shift towards much closer integration of data and statistics has taken place, in line with the Strategy for the Modernization of Statistics. This led to the establishment of a Data Coordination Group (DCG) with the mandate to ensure a strategic and technical coordination mechanism with recognized authority to make decisions of corporate relevance for data for statistics (e.g., big data) and statistics. The commitment for data and statistics integration is also embedded in the new Data and Statistics Quality Assurance Framework (SDQAF), which ensures a proper use of alternative data sources to fill data gaps and adherence to the highest quality standards of the resulting statistical products.

An additional corporate Standard on the Acquisition and use of non-statistical data sources for statistical purposes, has been developed to address quality aspects specific to different types of input data (big data classes), providing tailored recommendations and guidelines to statistical units. While focused on statistical purposes, this standard can be used to guide experimental statistics analyses, thus encouraging science and research. It also includes provisions on cases where FAO acquires data through a third party or partnerships, thus compensating for the lack of internal resources and promoting knowledge sharing. The Standards on Imputation and Quality indicators have also been revised to reflect the paradigm shift and include implications for the definition and computation of coverage errors and accuracy.

Lastly, the revised Quality Assessment and Planning Survey (QAPS) aiming to assess the compliance of FAO statistical activities with SDQAF principles was implemented in 2023. It provides insights about progress achieved on the quality at both process and output levels and highlights areas of improvement. Also, the systematic review of reference metadata against the recommendations of the revised Standard on Metadata dissemination contributes to improve the accessibility and clarity of FAO statistics.

This paper presents the tools and mechanisms put in place in FAO to ensure the production of high-quality statistical outputs while integrating non-traditional data sources into statistical processes.

Keywords: Quality Assurance Frameworks, Non-statistical data sources, Quality Assurance Mechanisms, Statistical Governance, Modernization.

1. Introduction

The Food and Agriculture Organization (FAO) is a specialized agency of the United Nations that leads international efforts to defeat hunger. Statistics is a core component of FAO's overall mandate as stated in the Article I of its Constitution: "the Organization shall collect, analyse, interpret and disseminate information on nutrition, food and agriculture". FAO also plays a central role in the international statistical community, being one of the key organizations of the International Statistical System (ISS), and a leading agency in the implementation of the 2030 Agenda for Sustainable Development¹. It is characterized by a decentralized data and statistical system operating under the coordination and oversight of a Chief Statistician, where the Statistics Division as well as other technical units are responsible for the production and dissemination of thematic statistics on several domains relevant to the Organization's mandate. Considering the large number of units and stakeholders involved in the statistical production process, enhanced governance and quality assurance mechanisms are vital elements to ensure the generation of data and statistical outputs satisfying highest quality standards and serving as global public good. In this framework, this paper discusses the institutional mechanisms and tools put in place by FAO to assure and monitor statistical quality and modernize the Organization's statistical system, including in the rapidly evolving data ecosystem.

2. Strategic repositioning towards an integrated governance mechanism and quality assurance framework for data and statistics

The establishment of corporate mechanisms for the coordination of FAO's statistical system and the quality assurance of its overall statistical programme is not new for the Organization. In 2012, the role of FAO Chief Statistician was created together with an Interdepartmental working group (IDWG) on statistics with the objective to improve inter-divisional coordination and cooperation on statistical programmes and promote corporate consistency and alignment in statistical practices within the Organization. In 2014, FAO adopted its first Quality Assurance Framework and, in the following years, a series of standards and procedures to improve the governance and quality of FAO statistical activities.

In 2019 however, an evaluation of FAO's statistical work led to several recommendations, including for the Organization to revise its statistics governance to ensure better coordination

¹ FAO is the custodian agency of 21 Sustainable Development Goal indicators and contributing to an additional 5 indicators.

and coherence of its statistical work and to accelerate actions to improve the quality of its data and IT infrastructure support. A corporate foresight exercise on how to address these recommendations led to an internal Strategy for the Modernization of FAO Statistics (FAO, 2020), which identified as priority actions the integration and improvement of the governance of FAO data and statistics to strengthen the oversight and enhance the relevance, coherence and quality standards of all its data and statistics, and the adoption of an innovative IT infrastructure and data management system to support the entire production cycle of FAO data and statistics. The implementation of these priorities and related outcomes are discussed below.

Integrating and improving the governance of FAO Data and Statistics

In 2022, the governance and coordination arrangements in place for FAO statistics were modified to capture the paradigm shift towards a much closer integration of data and statistics. The new internal coordination mechanism largely follows the recommendations of the UN Secretary-General's Data Strategy, whilst reflecting the critical role of data as an accelerator toward the delivery of the FAO Strategic Framework 2022-31 (FAO, 2021a). Under the new arrangements, an Executive Data Champion from FAO core management team was appointed and tasked to chair FAO Data Coordination Group (DCG). This DCG, which replaced the existing IDWG, oversees and is supported by a renewed overarching technical task force (DCG-T), chaired by FAO Chief Statistician, that convenes in plenary to address cross-cutting issues, yet also convenes in thematically focused working groups addressing statistics, big data and geospatial information. In fact, since May 2022, three DCG-T task forces were created with the specific mandate to tackle specific coordination and quality management gaps: a task force on statistical classifications and statistical metadata standards, a task force on integrated quality assurance framework and a task force on the land cover and land use statistics. The work of these task forces led to the development and endorsement of the new FAO Statistical and Data Quality Assurance Framework and new corporate standards to guide and better regulate the use of non-traditional data sources in the production of FAO data and statistics.

A Data Lab for Statistical Innovation was also created as a corporate mechanism to encourage, facilitate, and accelerate the use of alternative data sources and innovative methods to process unstructured data to fill existing data gaps and improve the quality of FAO statistics.

Strengthening FAO Quality Assurance functions

In addition to the changes in FAO's governance on data and statistics, a permanent and dedicated Data Quality Assurance Unit was created under FAO Statistics Division in January 2024 with the overall mandate to ensure that FAO produces and disseminates high-quality

data and statistics related to food and agriculture. This new unit was tasked with the following functions: (1) the development, review and implementation of FAO's Statistics and Data Quality Assurance Framework (SDQAF) and related corporate policies, statistical standards, tools and quality assurance procedures; (2) the quality monitoring and assessment of FAO data and statistics and, as necessary, the development and implementation of quality improvement plans; (3) the implementation of the accountability framework for FAO statistical activities; (4) the regular conduct of user consultations on FAO databases; and (5) the provision of capacity development on data and statistical quality assurance. While these functions were already embedded in those of FAO Chief Statistician and its office, the creation of a dedicated team and the recruitment of a senior statistician to lead it is expected to significantly strengthen and accelerate their delivery in the future.

The new FAO Data and Statistics Quality Assurance Framework (SDQAF)

Strengthened governance and better integration of data and statistics were also reflected in the new FAO Statistics and Data Quality Assurance Framework (SDQAF), adopted by the DCG in June 2023 (FAO, 2023a). The SDQAF represents an integrated and harmonized framework aiming to ensure that statistical and non-statistical data sources are used properly to fill existing gaps, and that resulting statistical and data outputs satisfy the highest quality standards and principles. The SDQAF is articulated around 16 principles that should be adhered to ensure the quality of the Organization's data and statistics (Figure 1 in Annex). They concern the institutional environment (principles 1-7), the implemented statistical processes (principles 8-11), and the produced statistical outputs (principles 12-16).

The new framework builds on the first FAO SQAF, extending its scope also to the use of non-traditional data for statistics, as well as to data-related concerns linked to the right to privacy, data protection, intellectual property rights and the need for transparency in data generation and acquisition. Among references taken into consideration are the Suggested framework for the quality of big data (UNECE, 2014), the recommended practices on the use of non-official sources in international statistics (CCSA, 2005) and FAO policies on data protection (FAO, 2021b) and intellectual property rights (FAO, 2023b). As a result, the new framework includes a new principle on suitable and trustworthy data sources and new key implementation modalities (previously referred to as best practices) in principles related to cooperation with data providers, data protection and statistical confidentiality, sound methodologies and appropriate statistical procedures, accessibility and clarity, and accuracy and reliability. The endorsement of the new framework by the DCG was followed by a thorough review of the main FAO quality assurance and assessment tools and mechanisms, as described in the following section.

3. Improved FAO Quality Compliance Mechanisms

With the adoption of the SDQAF in 2023, the FAO has also expanded and improved its quality compliance mechanism, which is composed of a series of tools and procedures to evaluate existing and new statistical processes and data sources to ensure their alignment to quality principles and best practices. First, the Quality Assessment and Planning Survey (QAPS) – a corporate self-assessment tool used to analyse and report on the quality of FAO data and statistics – was updated to reflect changes in the SDQAF and the implementation of the new policies on data protection and intellectual property protection. In the fourth quarter of 2023, nearly 100 FAO statistical processes, databases and information systems were assessed through the QAPS to, on the one hand, reporting on the quality of ongoing data and statistical processes at FAO, and, on the other hand, identifying areas for improvement. Second, the FAO is in the process of introducing more formal in-depth audits to complement the self-assessment produced with the QAPS, which will be implemented on a reduced number of key statistical outputs and databases. Auditing will be carried out by a team of internal and external auditors who will study statistical processes, supporting documentation and meet officers responsible for the process to further investigate any issue identified. A checklist is already in place to facilitate the completion of the task, the results of which will be summarized in a final report providing recommendations for quality improvement. Over the years, the implementation of the QAPS and informal ad-hoc audits has allowed identifying several quality gaps that have contributed to the development of the FAO series of statistical standards. Most of these standards, which were produced before the 2023 SDQAF, have been updated to better reflect quality practices for the integration of statistical and non-statistical data sources, as discussed in the next section, and a specific standard on the acquisition and use of big data has been developed.

Another quality assurance mechanism in place at FAO (called the Statistical Accountability Framework) entails the clearance by the Chief Statistician of new data collections and statistical activities, new dissemination platforms, any recruitment of statisticians and projects with significant data and statistical components undertaken by FAO. This process aims to ultimately ensure the relevance and quality of FAO procedures, processes and products. The underlying clearance mechanisms are being improved to consider FAO's evolving quality assurance and data-related frameworks, policies, standards and procedures.

Finally, in the context of an ongoing FAO large-scale dissemination project where all the Organization's statistical databases are gradually being migrated to the new Statistical Data Warehouse (SDW) using SDMX-native tools, a few quality compliance assessments, including

the validation of the source data and metadata and implementation of necessary adjustments in line with applicable SDMX and FAO Corporate standards, are systematically carried out beforehand.

4. Reflecting the integration and implications of using non-traditional data sources for statistical purposes in FAO standards

The FAO SDQAF has been complemented in 2023 by a dedicated statistical standard on the acquisition and use of non-statistical data sources for statistical purposes (e.g., filling data gaps, developing indirect estimation methods to generate more disaggregated statistics, etc.) which addresses specific quality aspects inherent to alternative data sources (FAO, 2023c).

As far as non-statistical data sources are concerned, a three-phase cycle, adapted from the UNECE Big Data Quality Framework (2014) and the Eurostat ESSnet Big Data project II deliverables on Methodology and Quality (2020), and consisting of Input phase, Throughput phase I and II and Output phase, is used (Figure 2 in Annex). FAO Standard, for instance, introduces some specificities compared to the ESS and UNECE frameworks. For instance, it recognizes that accessing raw data might not be feasible or cost-efficient for the Organization, and therefore proposes additional safeguards to assess and ensure the quality of data pre-processed by a third-party or the data owner/provider itself.

Compliance assessment checklist and appropriate agreement instruments

During the acquisition process, the Standards require certain pre-acquisition analytical work to be conducted starting from the input phase and even before the acquisition of input data. To this end, a process covering all data acquisition steps (identification, formulation, appraisal and clearance, and implementation) has been established (Figure 3 in Annex). Before final approval by designated entities and operationalization of the data acquisition, the process entails two types of assessment:

- ✓ *Assessment of the feasibility and cost-benefits of using the data when an agreement with the data owner/producer or a third party is needed:* depending on the type of partners (government entities, regional/international organizations vs. private sector or other non-state actors) and whether or not the agreement involves a financial contribution or other collaboration activities, the Standard provides recommendations on the appropriate instruments to use, including data sharing agreements (plus due diligence screening on the data provider/owner against a set of determined criteria), Memorandum of

Understanding (MoU) (plus application of the FAO's engagement framework² with private sector and non-state actors), with the aim to achieve a balance between benefits and potential risks arising from prospective engagements and preserving FAO's integrity, independence, and impartiality.

- ✓ *Assessment of the data source against cross-cutting considerations*³: a checklist, complemented with additional considerations associated to specific data sources, has been developed in this regard to ensure that the new data sources used by FAO are in line with the SDQAF principles, in particular Principle 8 on "Suitable and trustworthy data sources" and to facilitate their assessment against principles outlined in the Data Protection and the Intellectual Property Rights policies of FAO.

Provision of best practices on quality requirements specific to Big Data classes

The Standard provides guidance on quality aspects related to big data classes at Throughput phase I, when FAO is involved. It also emphasizes how crucial it is to use code reproducibility best practices while documenting every step of the data processing to ensure transparency of the overall process. When data is pre-processed by the data owner/provider, a documentation built on an optimal trade-off between public transparency, privacy and confidentiality, and industrial secrecy and intellectual property rights should be requested by FAO. The Throughput phase II involves the use of more traditional statistical procedures that are not specific to the data class.

Revision of other existing data and statistical standards as per the Standard's requirements at the Output phase

The adoption of the SDQAF and the Standard on the acquisition and use of non-statistical data sources for statistical purposes required the revision of several existing standards to ensure alignment and consistency between the various documents, in particular those targeting the output phase, where data and statistical outputs along with their corresponding metadata are disseminated. This concerns the following corporate statistical standards:

- ✓ *Standard on Observation status code list* (FAO, 2023d), revised to ensure alignment with SDMX observation status codes and to provide guidance on the assignment of observation status codes for derived variables and aggregates that are based on a mix of official and non-official values.

² Framework for Due Diligence and Risk Assessment/Management for Engagement (FRAME)

³ These include institutional and ethical, complexity and usability, cost and resource, and quality considerations

- ✓ *Standard on Quality Indicators for external users* (FAO, 2023e), revised to update the expression of a few quality indicators in line with the update of the standard on observation status. These include the percentage of observed (previously official) data points and the contribution of the imputed values to the final aggregate.
- ✓ *Standard on Imputation* (FAO, 2023f), revised to provide recommendations related to the use of statistical (machine) learning methods for imputation purposes, and guidelines/options to manage possible disputes with countries when using imputation to fill official data gaps.
- ✓ *Standard on Metadata dissemination of FAO statistical databases* (FAO, 2023g), updated to expand the scope and include information on the use of non-statistical data sources in the generation of statistical outputs, but also to ensure compliance with FAO Data Policies on Data Protection and Intellectual Property Rights.

In addition to the above-mentioned standards, new standards on metadata dissemination on geospatial information and on the dissemination of dashboards, data platforms and data visualization are being developed.

5 Other quality improvement initiatives

In terms of advancing methods and quality assurance best practices at the international level, FAO, as the secretariat of the [United Nations Committee of Experts on food security, agricultural and rural statistics](#) (UN-CEAG), is deeply committed to contribute to support the group in developing and documenting good practices and guidelines on concepts, methods and statistical standards for food security, sustainable agriculture and rural development statistics. Recognizing the specific strengths and weaknesses of agricultural statistics, the UN-CEAG 2020-2023 programme of work establish a line of work on the development of national quality assurance frameworks (NQAFs) for agricultural statistics to support national statistical agencies in assessing and improving the quality of statistics related to food and agriculture. As a result, the UN-CEAG produced and piloted test NQAFs focussing on three agriculture statistics subdomains: (a) crop and livestock production statistics; (b) statistics on producer prices of agriculture commodities; and (c) statistics on land use for agricultural purposes⁴. The structure of the developed self-assessment checklist was inspired by the International Monetary Fund data quality assurance framework (2012), and the 2019 edition of the United Nations National Quality Assurance Framework (UNDESA, 2019).

⁴ The checklist and scoring methodology proposed for the three framework is available on the FAO UN-CEAG webpage: <https://www.fao.org/about/ce-on-food-security-agricultural-rural-statistics/en/>

The proposed self-assessment tools can be used to evaluate compliance with both general quality principles and standards and guidelines for domain-specific statistics. They represent an excellent starting point for improving key elements of the national statistical production process in each subdomain, for designing statistical processes ex novo, and for producing new checklists covering other subdomains of agricultural statistics that have not yet been covered. Another initiative implemented jointly under the umbrella of the UN-CEAG and the UN Committee of Experts on Big Data and Data Science for Official Statistics (UN-CEBD) concerns the use of Earth observation data for agricultural statistics. With the objective of supporting countries through the provision of methods, tools and trainings on the use of EO data for crop acreage and crop yield estimation and mapping, the UN-CEAG/UN-CEBD joint task team has done work on: (a) the optimization of in situ field survey design; (b) the efficient preprocessing of satellite imagery; (c) the extraction of phenospectral features; (d) the use of different classification algorithms; (e) the validation of results; and (f) the creation of training curricula on the use of EO data for statistics. In 2022 and 2023, the joint task team worked on a series of key use cases in countries to optimize the design of field surveys and data georeferencing protocols, with a view to increasing the accuracy of crop type maps developed using in situ data. It also collaborated closely with the Big Data Hubs to further disseminate methods and best practices in this area.

6 Conclusions

FAO data and statistical system operates in a continuously evolving environment, characterized by the proliferation of new data sources and technologies, and the constant emergence of new demands and priorities. This comes with enormous opportunities to increase quality of statistical products in terms of their granularity, timeliness, and relevance, but also with several challenges that need to be dealt with in an integrated data and statistics quality assurance framework. The present paper discussed the new governance structure and quality assurance framework and mechanisms that FAO has put in place to implement its statistical modernization strategy and ensure statistical processes and outputs satisfying the highest quality standards. It is expected that progress made in artificial intelligence will continue to shape future work in this area as they pose additional challenges in terms of methods, ethics, data quality and transparency that the world is only starting to grasp. FAO White Paper on Responsible AI (2024) is proposing additional actions to the ones described in this document, such as the adoption of a framework for responsible AI and additional internal control mechanisms which will have an impact on the use of AI for statistical purposes.

References

- CCSA (Committee for the Coordination of Statistical Activities) (2015). [Recommended Practices on the Use of Non-Official Sources in International Statistics](#).
- Eurostat (2020). [Revised Version of the Quality Guidelines for the Acquisition and Usage of Big Data](#). Deliverable of the ESSnet Big Data II project.
- FAO (2014). FAO Statistics Quality Assurance Framework. Rome, Italy.
- FAO (2020). [Strategy for the Modernization of FAO statistics](#). Rome, Italy.
- FAO (2021a). [Strategic Framework 2022-31](#). Rome, Italy
- FAO (2021b). [Data Protection Policy](#). Rome, Italy.
- FAO (2023a). [FAO Statistics and Data Quality Assurance Framework](#). Rome, Italy.
- FAO (2023b). [FAO Policy on Intellectual Property Rights](#). Rome, Italy.
- FAO (2023c). [Statistical Standard Series: Acquisition and use of non-statistical data sources \(including Big Data\) for statistical purpose](#). Rome, Italy. Internal document.
- FAO (2023d). [Statistical Standard Series: Observation Status Code List](#). Rome, Italy.
- FAO (2023e). [Statistical Standard Series: Quality Indicators for External Users](#). Rome, Italy.
- FAO (2023f). [Statistical Standard Series: Imputation](#). Rome, Italy.
- FAO (2023g). [Statistical Standard Series: Metadata Dissemination for FAO Statistical Databases](#). Rome, Italy.
- FAO (2023h). [Tracking progress on food and agriculture-related SDG indicators 2023](#). Rome, Italy.
- FAO (2024). FAO White paper on responsible AI: Leveraging Artificial Intelligence for Food and Agriculture. Rome. Internal document.
- IMF (International Monetary Fund) (2012). [Data Quality Assessment Framework](#) – Generic Framework.
- UNECE (United Nations Economic Commission for Europe) (2014). [A suggested framework for the quality of big data](#). Deliverables of the UNECE Big Data Quality Task Team. Geneva, Switzerland.
- UNDESA (United Nations Department of Economic and Social Affairs) (2019). [United Nations National Quality Assurance Frameworks for Official Statistics](#). New York, USA
- UNSC (United Nations Statistical Commission) (2022a). [Report of the Food and Agriculture Organization of the United Nations on recent developments in agricultural and rural statistics](#). New York, USA.
- UNSC (United Nations Statistical Commission) (2022b). [Trusted methods: Lessons Learned and Recommendations from Select Earth Observation](#). New York, USA.
- UNSC (United Nations Statistical Commission) (2024). [Report of the Food and Agriculture Organization of the United Nations on recent developments in agricultural and rural statistics](#). New York, USA.

Annex.

Figure 1: Vision, principles, quality dimensions, and compliance mechanisms of the FAO SDQAF



Figure 2: FAO's business process involving the use of non-statistical data sources for statistical purpose

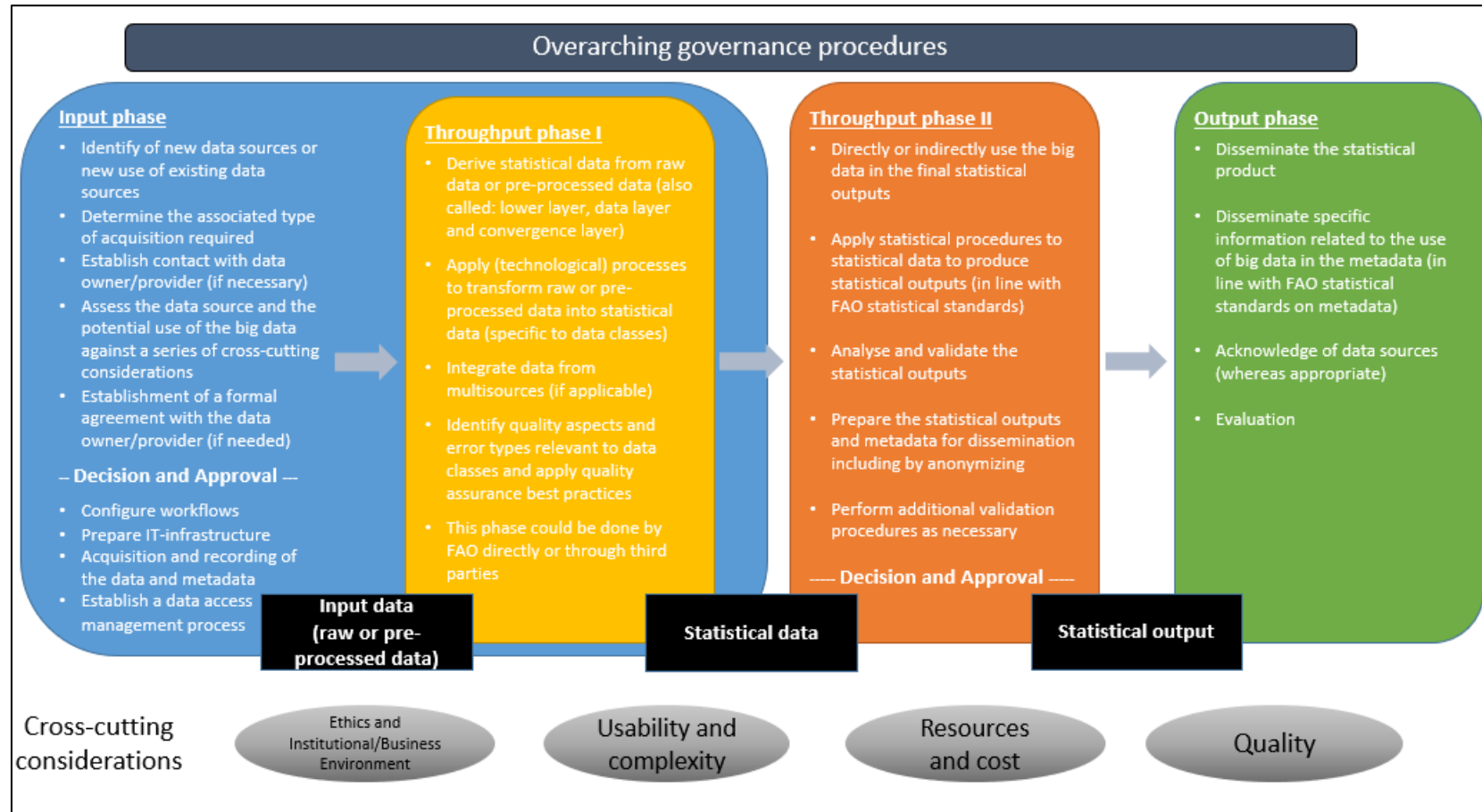


Figure 3: Data acquisition process

