

Distr.  
GENERAL

Working Paper  
... April 2013

ENGLISH ONLY

**UNITED NATIONS  
ECONOMIC COMMISSION FOR EUROPE (ECE)  
CONFERENCE OF EUROPEAN STATISTICIANS**

**EUROPEAN COMMISSION  
STATISTICAL OFFICE OF THE  
EUROPEAN UNION (EUROSTAT)**

**ORGANISATION FOR ECONOMIC COOPERATION  
AND DEVELOPMENT (OECD)  
STATISTICS DIRECTORATE**

**Meeting on the Management of Statistical Information Systems (MSIS 2013)**  
(Paris, France, and Bangkok, Thailand, 23-25 April 2013)

Topic (ii): Streamlining statistical production

## **Meeting the Future Demands of a Statistical Organization**

Prepared by Laurent Meister, Statistics Department, International Monetary Fund (IMF)<sup>1</sup>

### **I. Introduction**

1. The paper discusses essential components of a streamlining strategy for statistical data processes with specific examples from the initiative of "Standardize, Streamline, and Automate"—a possible option that could be considered for implementation within the IMF Statistics Department. The primary topics to be covered include: Standards - with respect to business processes, their underlying components and business roles; Technology - desirable system and user-level functionalities that enhance production efficiencies; Organization – a desirable framework to support production activities and to maintain the efficiencies gained over time; and Validation Methodologies – efficiencies possible within the “identification, investigation, decision and action phases” that are part of the validation process.

### **II. Business Context**

2. In the wake of the 2008 financial crisis, the international community identified important data gaps that needed to be filled in order to create a global information system suitable for monitoring global financial and nonfinancial flows and positions comprehensively, as well as monitoring the ongoing impact of inter-connectedness. In particular, the report *The Financial Crisis and Information Gaps* prepared by the staffs of the Financial Stability Board (FSB) Secretariat and the International Monetary Fund (IMF) and endorsed by the Group of Twenty (G-20) Finance Ministers and Central Bank Governors in November 2009, identified areas of data collection and processing that would need

---

<sup>1</sup> The views expressed herein are those of the authors and should not be attributed to the IMF, its Executive Board, or its management.

to be addressed by the IMF and other international organizations. It is anticipated that, primarily in response to the G-20 Data Gaps Initiative, which the volume of data processed by the IMF Statistics Department would increase by a factor of four over the next five years. The increased demand is expected to have to be accommodated within existing budgetary resources.

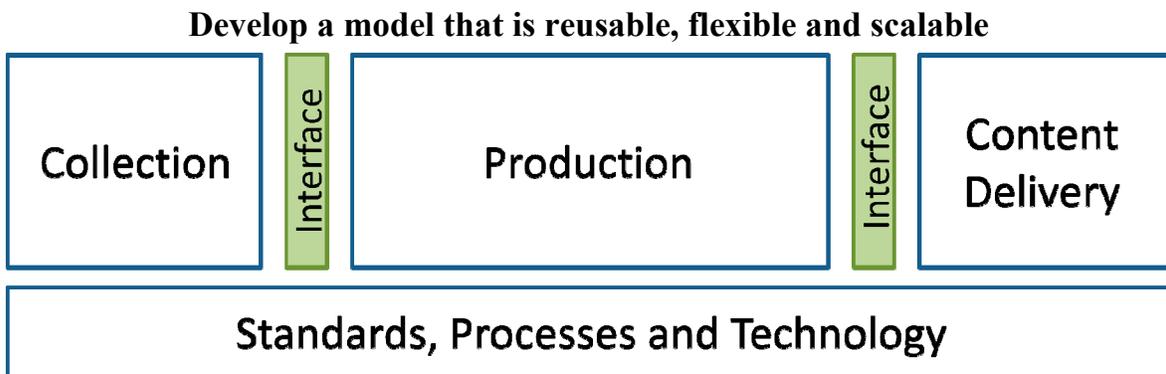
### III. Overview

3. The following is an overview of the specific objectives that the IMF Statistics Department (STA) plans to achieve to respond to the impending business context together with a suggested approach for achieving these objectives.

- Meet the rapidly increasing demands for more data and metadata products
  - ✓ Develop a model that is reusable, flexible and scalable
- Increase the timeliness of data and metadata delivery
  - ✓ Increase efficiency of data and metadata collection and processing
- Reduce the incidence of data and metadata errors
  - ✓ Increase the quality and volume of data and metadata validation performed per product
- Increase process and content transparency
  - ✓ Support end-to-end visibility, traceability and accessibility of production processes and production contents
- Enable content delivery to meet the increasing diversity of consumer needs
  - ✓ Develop a portfolio of multiple delivery methods
- Maintain strong relations with users
  - ✓ Develop and execute communications strategies with key user groups

### IV. Objectives and Goals

#### A. Meet rapidly increasing demands for more data and metadata products



4. Our analysis of the current STA statistical production systems concluded that it is possible to find a generalized approach that could be reused across statistical products. The scalability, reusability

and flexibility needed to quickly create multiple new statistical products could be achieved through workforce specialization and through the adoption of the customizable process template.

(a) Business Implications

In broad agreement with the Generic Statistical Information Model (GSIM) recommendations, it was found that statistical activity is focused on three primary domains: collection, production (data processing) and content delivery and that each domain could be standardized across statistical products. The new business processes would be aligned to these centers of activity and these, in turn, would be the drivers for the organizational and technology support.

The current “cottage industry”, where customized business processes were developed for each statistical product, would migrate towards standardized business processes.

(b) Organizational Implications

(i) Collection

There are a wide variety of collection methods that could be shared across all Statistical Products. A specialized team, with a broad knowledge of the multiple formats and collection tools, dedicated to all collection activities could effectively increase the scalability of these operations.

(ii) Production – Data Processing

A common production process template has been identified and guidelines developed for its implementation across statistical products. The common standards underlying the process template would meet the most significant validation needs of various STA’s production processes, while maintaining the flexibility needed for product-specific customizations. The implementation of these standards would facilitate the scalable development and management of new and existing statistical products within STA. A specialized team dedicated to all data processing activities combined with standardized production environments could effectively increase the scalability of these operations.

Please refer to the Appendix for a more detailed exposition of the Standard Production Process Template.

(iii) Content Delivery

There are a finite number of content delivery methods shared across all data and metadata products. A specialized team dedicated to all content delivery activities could effectively increase the scalability of these operations.

(iv) Standards, Processes and Technology

A team specialized in the overall design and implementation of the standards, processes, and technologies supporting the operations related to statistical products could help ensure that investments made in standardization are realized, sustained, and enhanced over time.

Standards: This team would be responsible for the design and ongoing enhancement of all content, process and technology standards supporting the statistical product operations and could also be responsible for the adoption of the standards in production processes.

Technology: This team would be responsible for ongoing enhancements to technologies supporting the statistical product operations and could also be responsible for organizational adherence to the “best practices” in the use of the supporting technology, to maximize efficiencies.

Operations: This team would be responsible for the initial statistical product setup and the ongoing compliance of all statistical products to the standards.

(c) Technology Implications

Statistical production processes would move away from product-specific, often home-grown, implementations and tools to a common toolset within each primary domain (collection, production, content delivery), reducing the need for staff training on multiple technologies and increasing their scalability across statistical products as production operations grow.

Standardized technology interfaces between the primary domains could provide avenues for independent growth of each domain.

Each domain would be able to have independent functional growth. For example: (1) As new collection methods and formats are created, they would immediately be able to feed existing production processes; (2) When new production processes are implemented for new statistical products, they would automatically be able to reuse existing collection and content delivery processes.

Each domain would be able to independently migrate to newer technologies, whenever it is ready to do so. As long as the common interface is maintained, within a domain, it would be possible to migrate from older, less functional technologies to newer ones without disrupting work performed in the other domains.

**B. Increase the timeliness of data and metadata delivery**

*Increase efficiency of data and metadata collection and processing*

5. Efficiency gains could be achieved through workforce specialization and through the adoption of technologies that support automation.

6. Timeliness of content delivery is dependent on three factors:

(a) The time for the originating institution to compile the data and metadata.

This will not be addressed within the context of this paper. Support could be offered through technical assistance and training programs.

(b) The time for the originating institution to communicate the contents to the IMF Statistics Department.

Collection: A specialized team could not only be efficient in the administration of collection activities, but could also be dedicated to help partner institutions submit in an increasingly automated manner, as their technical maturity evolves, with the goal of bringing them to fully automated, most timely, machine-to-machine content exchanges.

(c) The time for the IMF Statistics Department to process and deliver the contents.

(i) Production

The Standard Production Process Template offers a highly automated production process where intervention by support staff or statisticians should only be required when expert judgment and decisions are required. Maximum efficiency requires the use of workflow technologies that could automate the routing of tasks to the relevant staff based on diagnostic flags. Please refer to Appendix for a schematic overview of a workflow based system.

(ii) Content Delivery

Content Delivery should be fully automated unless there is a business need to control delivery, e.g. to coincide with a press release. Standardized production outputs should feed automatically on a scheduled or business-user controlled basis to the content delivery media.

## **C. Reduce the incidence of data and metadata errors**

### ***Increase the quality and volume of data and metadata validation performed per product***

7. The increase in the quality and volume of data and metadata validation performed would be achieved through a standardized approach to validation across statistical product production processes and through the adoption of technologies that support advanced validation functionality.

8. Production: Content validation is the responsibility of production processes. The increase in the effectiveness of validation operations could be achieved through the following methods:

(a) Adoption of standardized methodologies for validation: ensure the consistency and comprehensiveness of validation operations across statistical products

## (b) Adoption of comprehensive validation functionality

- (i) Includes the adoption of technologies that would allow business users to specify and automate the following types of validation tests:

Intra-series (time series characteristics), inter-series (relations between time series), cross-dimension (e.g. cross-country/cross-concept/reported vs. published consistency tests), cross-database (e.g. cross-sector consistency tests), metadata (e.g. missing content, cross-metadata consistency), data-metadata consistency (e.g. data is equal or close to 100 at the identified base year time period), ad-hoc formulas, etc.

- (ii) Provides for high performance of validation operations

The adoption of technologies capable of automating and executing large volumes of validation tests, to cover all needed validation scenarios.

- (iii) Supports efficient analysis of validation results

Support for the prioritization and aggregation of validation results, to guide reviewers through results in order of importance and to display results at the appropriate level of detail for all users (managers, statisticians, support staff).

Integration of data and metadata from multiple sources to allow reviewers to have at their fingertips all the relevant related information that they would need to make informed decisions.

Support for presentation formats that display results in a highly visual and immediate manner. These should include the following: time series views, OLAP views, charts, visualizations.

- (iv) Efficient consequent actions

Support within the validation environment for ad hoc data or metadata corrections, the ability to document decisions and findings for future reviews and support for automated actions (e.g. workflow) based on diagnostic results.

#### D. **Increase Process and Content transparency**

*Support end-to-end visibility, traceability and accessibility of processes and contents*

9. End-to-end visibility, traceability and accessibility of processes and contents would be achieved through the process standards and the adoption of technologies that support these functionalities.

(a) End-to-End visibility

(i) Content

Internal consumers of the statistical products would be able to access, at their own choosing, data at each significant processing milestone: input, submission signoff, database pre-publication signoff, database publication signoff.

At all times, processing staff would be able to access diagnostics on the quality of the in-process and published data.

(ii) Processes

At all times, users would be able to know the status of any submission.

Consumers would be able to subscribe to notifications of process events and outcomes; e.g., consumers may choose to receive notifications when data for their country of interest have been updated.

(b) End-to-end traceability

(i) Content traceability requires substantial technological support. It would allow users to trace data transformations from:

Source to destination: this would support performing impact analysis of any input or transformation made to source data on published outputs.

Destination to source: to allow users to see all the inputs, transformations and manual actions used to produce a published time series or observation.

(ii) Process traceability would allow users to analyze detailed logs of all process actions and events.

(c) Accessibility

(i) Consumers without any specialized software would be able to access contents and process information over the web.

(ii) Consumers with specialized software would be able to store contents locally and work with them off-line.

**E. Enable content delivery to meet the increasing diversity of consumer needs**

*Develop a portfolio of multiple delivery methods*

10. For statistical products to be effective, they must be delivered in ways that meet the growing needs of users. A content delivery team could be dedicated to developing a variety of automated processes to publish statistical product outputs. At a minimum, the technology would support the delivery of contents in raw formats for research or as prepared analytical presentations (to tell a story) in an increasingly interactive way, on an on-demand basis across multiple media such as the web, web services, mobile and print publications.

**F. Maintain strong relations with users**

*Develop and execute communications strategies with key user groups*

11. Each domain team (collection, production, content delivery) could be responsible for supporting their respective user base and for collecting feedback to continually assess the relevance and effectiveness of their services. The assessments would allow domains to remain responsive to the evolving needs of the various user groups.

(a) Collection

A team could support and maintain relations with those responsible for the communication of the collected data and metadata and could collect feedback to continually assess the relevance and effectiveness of the collection methods offered.

(b) Production

A team could maintain relations with those responsible for the contents collected and could ensure that all input anomalies are either corrected or properly documented, in collaboration with them.

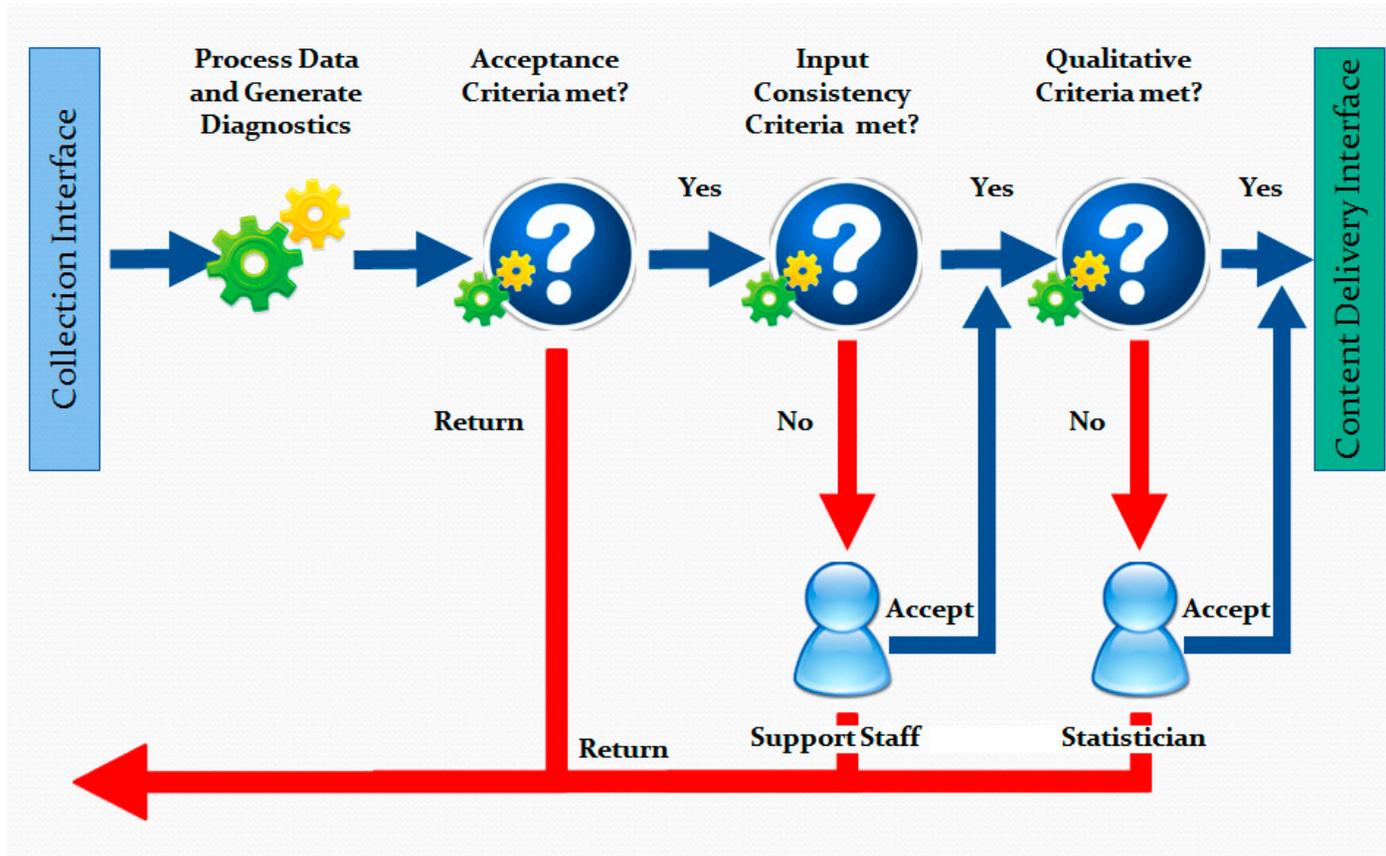
(c) Content Delivery

A team could support the users who consume delivered content could collect feedback to continually assess the relevance and effectiveness of the Statistical Products published.

(d) Standards, Processes, and Technology

A team could support and maintain relations with the collection, production, and content delivery teams and could collect feedback to continually assess the relevance and effectiveness of the standards, processes, and technology supporting the statistical products.

Standard Production Process Templates



The process model is customizable and demonstrates the possibilities of automation. On a per Statistical Product basis, one or more steps in the process may not be relevant and would be omitted, for example for certain statistical products no automation of qualitative criteria is possible and thus Statisticians would receive every submission for review. The process initiates with the complete processing of inputs and the creation of all diagnostics so that all information needed to make decisions would be present before any user performs a review. The workflow may then use key diagnostics to make an automated decision on the need to include support staff and/or statisticians into the review of the inputs and derived outputs. The outcomes of the automated and user reviews would always end either in the acceptance or return of the submitted inputs and their derived outputs.