منظمة
الأغذية والزراعة
للأمم المتحدة

联 合 国
粮 食 及
农 业 组 织

Food and Agriculture
Organization of the
United Nations

Organisation des
Nations Unies pour
l'alimentation et
l'agriculture

Продовольственная и
сельскохозяйственная
организация
Объединенных Наций

Organización de las
Naciones Unidas para la
Agricultura y la
Alimentación

AFRICAN COMMISSION ON AGRICULTURAL STATISTICS

Twenty-eighth Session

Pretoria, South Africa: 4 – 8 December 2023

AGENDA ITEM 9

New FAO Data dissemination platform: FAODATA Explorer - Beta Version

NAME OF AUTHORS
Aymen Charef ; Valerie Bizier

ORGANIZATION
Office of the Chief Statistician, FAO

SUMMARY

This paper presents an overview of FAO's modernization project related to the establishment of an integrated Statistical Data Warehouse (SDW) and FAODATA Explorer, FAO's new centralized dissemination platform. It describes the context and objectives of this project, its IT infrastructure and the workflow that supported the release of FAODATA Explorer beta version on September 15th, 2023, containing FAO SDG Data Domain (i.e. the data series for the 21 SDG indicators under FAO custodianship). The FAODATA Explorer is still under development and will gradually be populated with existing FAO statistics on food, agriculture, nutrition, fisheries and aquaculture currently disseminated through FAOSTAT, FishSTAT and other platforms.

The present paper also depicts how FAO harnessed the internationally agreed SDMX standard and related information model and tools to enable a fast pace of modernization and to bring efficiencies in FAO business processes. SDMX was used to standardize data and metadata exchange and dissemination to increase the quality of statistics disseminated by FAO, in compliance with the pillars of FAO Data and Statistics Quality Assurance Framework.

Members are invited to take note of the paper and express views on this FAO's modernization project, in particular on current functionalities of the FAODATA Explorer beta platform and the next steps of the project.

1. Introduction

Statistics is fundamental in FAO's work, supporting evidence-based decision-making of national organizations and the international community enabling progress monitoring towards national and international development goals. Article 1 of the FAO Constitution stipulates that *"The Organization shall collect, analyze, interpret and disseminate information relating to nutrition, food and agriculture."* where agriculture is meant to include fishery and forestry activities. FAO is acknowledged as an authoritative source of data and statistics at national and international levels in the food and agriculture sectors, which comprises, among others, data on crops,

livestock, fisheries, forestry, land and water resources and use, food and nutrition security, climate change and environmental impact of agricultural activities, as well as rural development.

FAO's statistical processes cover the entire production cycle, involving the collection, validation, compilation, production, analysis, and dissemination of statistics. FAO relies on a statistical production system and Information Technology (IT) tools to keep up with the internal processes and information needs of internal and external users, illustrating the organization ability to adapt to new developments in society, knowledge and technology and changing policy priorities.

The independent evaluation of FAO's Work in Statistics (FAO, 2020) concluded that "FAO needs to accelerate actions to improve the quality of its data and IT infrastructure support". To this end, it recommended that "An integrated statistical quality management system should be put in place and enforced by FAO, covering all statistical activities, to ensure full adherence to existing and new internationally accepted levels of statistical standards and norms" and "Procedures and information technology tools should be implemented by FAO for (statistics) input, processing and output, standardized within an overall data-warehousing strategy, to improve the quality of its data".

As a response, FAO developed a Modernization strategy for its statistics and a proposal for improved governance of FAO statistical activities was presented to the 132nd Session of the Programme Committee (FAO, 2021). The proposal is articulated across four cross-cutting priority action areas and one of the priority areas is to "Improve the quality of the IT infrastructure supporting data and statistics work".

Based on this proposal, the feedback received from member countries and inputs from multiple departments and divisions, the project "Modernization and integration of the FAO statistical system" was launched in 2022, is due to be finalized by 31 July 2024 and is supported by the Capital Expenditure (CapEX) fund.
The project identifies two components related to FAO Statistical IT infrastructure: the revamp of the Statistical Working System (SWS), as well as the setup of a Statistical Data Warehouse (SDW) and the integrated data dissemination platform FAODATA Explorer:

- On the one side, the SWS, consists of FAO internal system to support the design, collection, processing, and aggregation phases of FAO statistical value chain, will be enhanced. It will be modernized by replacing obsolete modules of the original software development framework and introducing new and innovative functionalities.

- On the other side, the project will implement the validation and dissemination phases through deployment of a centralized SDW and FAODATA Explorer, FAO's new integrated dissemination platform. The SDW will be fully integrated with the SWS, automatically harvesting its outputs and will also gather data from other production systems. The integrated data will be disseminated through an interface platform and open new channels and services of data exchange with national and international statistical organizations, as well as research institutions and the global user community.

The present paper focuses on the deployment of the SDW and the FAODATA Explorer. The first two sections describe the information model and the selected IT tools based on the Statistical Data and Metadata Exchange (SDMX) standard that were adopted by FAO for this project. The third section describes the role of the project in leveraging SDMX best practices to support FAO's endeavours of implementing its Statistics and Data Quality Assurance Framework (SDQAF) (FAO, 2023). The fourth section reports the workflow and processes used for the dissemination and reporting of SDG data. The final section presents the next steps of the project. Finally, the paper invites AFCAS members take note of this report and express views on the FAO's SDW and FAODATA Explorer project, in particular on the current functionalities of the beta platform and the next steps of the project.

2. Scope and objectives of the SDW and FAODATA Explorer

2.1. Centralization and integration
The overall objective of the setup of SDW and dissemination platform FAODATA Explorer is to (1) respond to users and evaluators requests of FAO data and statistics (2) integrate data from more disparate sources into a cost-effective central source and (3) expand FAO's outreach of statistical data thereby serving an increasing number of data users and providing solid evidence to respond to the main policy demands.

The first phase of the project (2021-2024) covers the migration of FAO data currently disseminated through FAO SDG data portal, FAOSTAT and FishSTAT. As first milestone, the FAODATA Explorer platform was made public on 15th September 2023 with the SDG data series of the 21 indicators under FAO custodianship, also disseminated in FAOSTAT and the UN Global SDG Database. FAODATA Explorer is currently a beta version as it is under development. It will be gradually populated with existing FAO statistics on food, agriculture, nutrition, fisheries and aquaculture currently disseminated through FAOSTAT and FishSTAT.

The consecutive phases aim to integrate other relevant FAO data assets currently disseminated through a dozen of statistical and data dissemination platforms, as well as the new databases that will be created in the future.

2.2. Data quality assurance

The deployment of SDW and FAODATA Explorer supports FAO's efforts to harmonize and standardize its data and metadata in compliance with the FAO SDQAF (FAO, 2023). In particular, it supports the development and implementation of corporate statistical standards, definitions and classifications, as well as methodological guidelines for the harmonization of the procedures implemented by the different statistical processes, all aimed at achieving a more integrated data and statistical system.

For this project, the work of harmonization of data and metadata models will be pursued applying best practices and harnessing the SDMX standard and related tools to deliver on business objectives, such as making data more accessible, interoperable, comparable and coherent.

In compliance with SDQAF, harmonizing concepts and terminology that are common to many statistical domains is making data processing more efficient through automation or delivering on data quality objectives. FAO data producers will take necessary time to streamline the workflow of statistical data from data sources to FAODATA Explorer for an improved internal data sharing and data access for external data users.

Furthermore, FAO will undertake user consultations to monitor the implementation of FAO SDQAF and to assess users' perception of the FAO DATA Explorer data with regards to FAO quality dimensions. It is already expected that overall quality of FAO statistical outputs will be enhanced, for instance:
- coherence, comparability: data, structural and reference metadata will be standardized, harmonized and made available in comparable format.
- accessibility, clarity: FAO data and statistics will be available through one point of access/dissemination platform and accompanied by improved metadata.
- timeliness and punctuality: FAO Data will be available as soon as it is produced using streamlined, more effective and effortless workflows.

These quality dimensions can be directly measured using ad-hoc methods or quality indicators.

3. Information model

The statistical workflow and information model used for data dissemination and exchange through FAODATA Explorer are entirely based on the SDMX standard, which was endorsed by the United Nations Statistical Commission in 2008 as a preferred standard for data exchange and was approved as an ISO standard (ISO/IS 17369:2013)[1].

The SDMX information model was leveraged to enable modelling and standardization of statistical data, metadata and the processes of data dissemination and exchange. The underlaying information model also identifies objects and their relationships and allows central management and standard open access.

The SDMX-based tools, presented in the next section, are the cornerstone to streamline the current management of data dissemination, standardize and automate processes, reduce the maintenance costs of data cubes and derived data.

The SDMX guidelines and best practices were used and adapted to the FAO context. The SDMX Glossary Version 2.1 (SDMX, 2020) contains the common SDMX terminology of concepts and related definitions used

---

[1]SDMX ISO standard : https://sdmx.org/?sdmx_news=un-statistical-commission-sdmx-is-preferred-standard

in structural and reference metadata of international organizations and national data-producing agencies. The most used terms in this document are:

- Structural metadata comprises the statistical concepts, code lists and data structure definitions (DSDs) which provide the basis for the data representation of each dataset.

- Data Structure Definition defines the dimensions, attributes and measures (representing the phenomenon to be measured) in a dataset and associates them with common code lists and concepts used in representing data in a particular subject-matter domain.

- Reference metadata consists of the information about quality descriptions, process descriptions, methodological descriptions and administrative descriptions. Reference metadata are described in a standard way using the Metadata Structure Definition (MSD).

4. IT infrastructure

Rather than developing the SDW and dissemination platform starting from scratch, FAO selected an off-the-shelf open-source scalable platform, called dotStatSuite. It is an SDMX native platform, and its development strategy is guided by the Statistical Information System Collaboration Community (SIS-CC)[2] according to the plans set by the community priorities[3]. The SIS-CC is composed of national and international statistical organizations that share common challenges and goals of promoting and adopting statistical processes and SDMX-based tools that facilitates data exchange to reduce duplications of efforts and implementation costs.

The dotStatSuite platform is comprised of three independent but interoperable components: dotStatCore, dotStatData Lifecycle Manager, and dotStat Data Explorer (DE). The last component is the modular front-end interface visible to external users as FAODATA Explorer, whereas the first two components are used in the back-end interface as described below.

4.1. Front-End

The front-end offers a consistent graphical interface that comprises a homepage for faceted search by categories (e.g. Area, Domain, etc) or free text search using a SolR-based search engine (Figure 1).

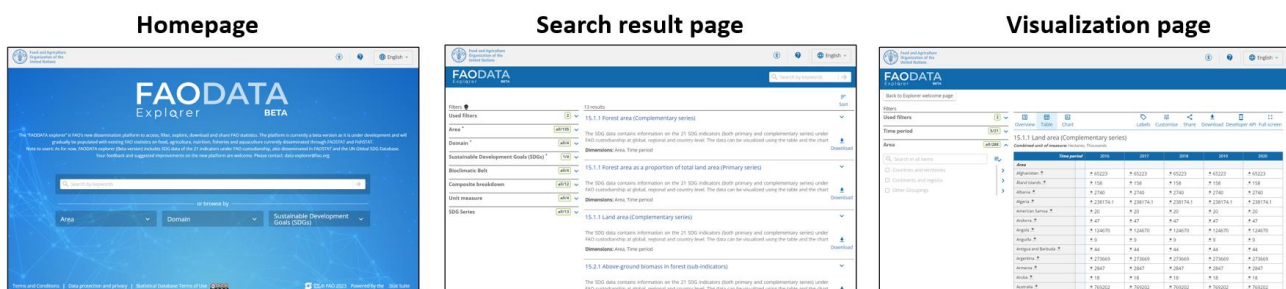| Homepage	Search result page	Visualization page

Figure 1: FAODATA Explorer as FAO-branded version of the open-source dotStatSuite Data Explorer.

The search result page and visualization page allow refining the search through filters and downloading the data in multiple formats (XLXS, CSV, SDMX-ML). It allows previewing data prior to download, using time series, bar charts, scatter plots, symbols and maps. For each dataset, an SDMX-based reference metadata (in CSV format) is attached according to FAO Metadata Structure Definition contained in the Statistical Standard on Metadata Dissemination for FAO statistical databases (FAO, 2023).

FAODATA Explorer also offers an SDMX 2.1 API for sharing data and structural metadata.

---

[2] Statistical Information System Collaboration Community: https://siscc.org/who-we-are/
[3] Flight planner: https://sis-cc.gitlab.io/dotstatsuite-documentation/about/product-overview/#flight-planner
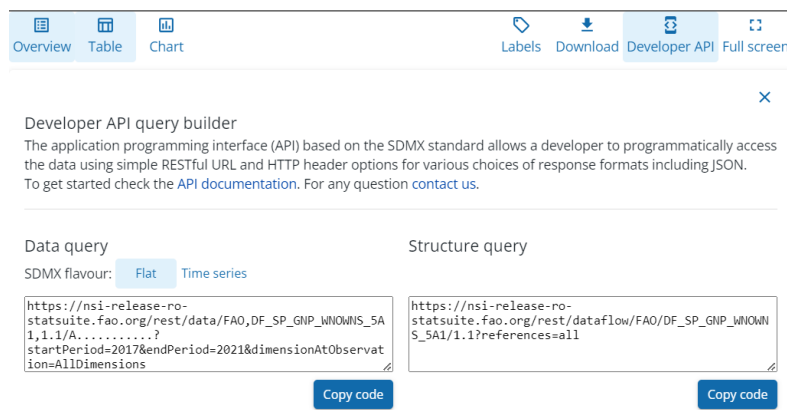
Figure 2: Documented API for sharing data and structural metadata.

4.2. Back-End

Besides the dotStatSuite components (dotStatCore and dotStatDLM) used in the back-end, FAO also used the SDMX native tool Meta and Data Manager of the SDW to create and submit structural metadata and build a dissemination and reporting SDMX compliant database (ISTAT, 2022). This open-source tool is developed and maintained by the Italian Institute of Statistics (ISTAT).

4.3. Feeding other IT platforms

The SDW will provide a centralized location of FAO statistics that spans diversification of the statistical data products and services. It will be the data source to feed modern interfaces, catalogues, SDMX registry and data visualization functionalities to improve users' experience and their perceptions of FAO as an innovative organization. For instance, SDG data currently disseminated in FAODATA Explorer is feeding the interactive visualizations, made using Tableau software, which are available in the SDG data portal[4].

The SDW and FAODATA Explorer will also open new channels and services of data exchange with national and international statistical organizations and offer the dissemination of interoperable structured data, concepts and classifications that can be harvested through APIs, according to FAIR principles (Findability, Accessibility, Interoperability, and Reuse) to global users, UN agencies, UN global data portal and UNSD databases (i.e. SDG Global Database). Furthermore, the organization will be able to exchange data with SDMX users in national and international organizations.

5. Dissemination and reporting of SDG data

FAO is as custodian agency for 21 SDG indicators for the implementation of the 2030 Agenda for Sustainable Development. The organization is responsible for the design and implementation of these indicators' methodologies, in agreement with Member Nations; and for the compilation, harmonization, validation and dissemination of related data and metadata, which are reported to the SDG Global Database hosted by the United Nations Statistics Division (UNSD), as approved by the United Nations Statistical Commission.

For this purpose, an information model was built for the SDG data integration, dissemination and reporting based on the global SDG DSD that was released by the Working Group on SDMX under the Inter-Agency Expert Group on SDGs (UNSD, 2019) to establish a uniform structure and concept definitions.

- For dissemination, the FAO SDG DSD included FAO specific codes and attributes.

- For reporting, the UNSD used the API to pull the data directly into their system of the global SDG database.

Figure 6 represents a generic workflow that is applicable for dissemination and reporting using the SDMX tools and following these main processes. This step-by-step workflow was implemented for the SDG data resulting in dissemination of sixty datasets in FAODATA Explorer and several automatic transmissions of SDG data. SDG data quality and consistency has substantially improved thanks to the steps of quality assurance and validation against the FAO SDG DSD.

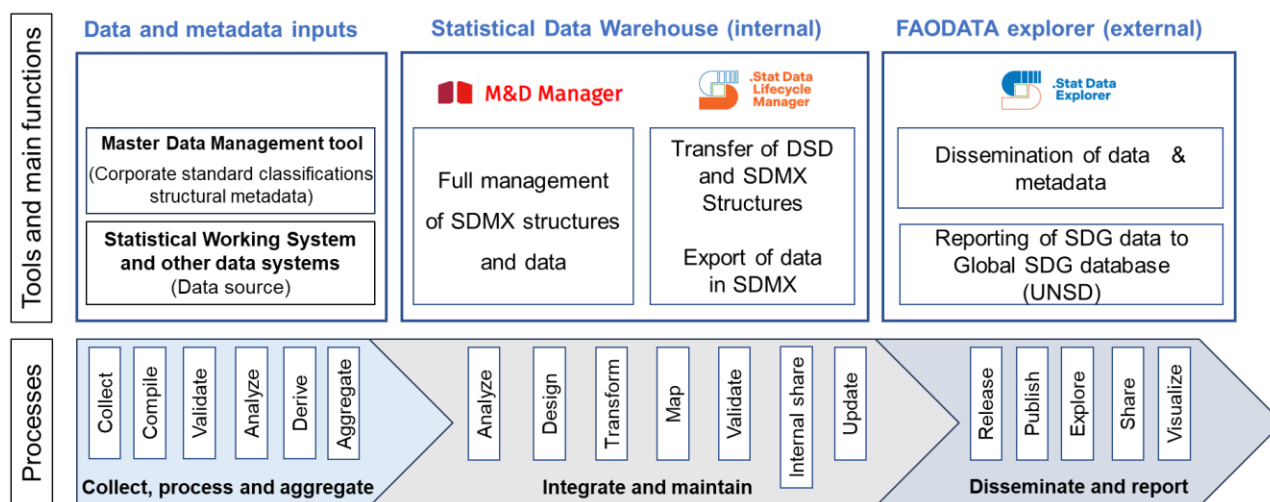[4] FAO SDG data portal: https://www.fao.org/sustainable-development-goals-data-portal/data/

Figure 6: Generic workflow for data dissemination in FAODATA Explorer and reporting to UNSD.

6. Next Steps

The activities under the first phase of the project will continue the modelling and integration of the bulk of FAOSTAT and FishSTAT data to be released by the end of next year. FAO is currently seeking feedback from FAO internal and external data users on the current and upcoming versions of the FAODATA Explorer Beta Version to ensure the platform has the best user-driven approach for its visualization and functionalities. Feedback can be provided by email at data-explorer@fao.org.

The next phase will focus on the integration of other relevant FAO data assets currently disseminated through a dozen of statistical and data platforms, as well as the new databases that will be created in the future.

The migration will be prioritized based on the importance of the data assets for users and the complexity of the data modelling and integration into the SDW, so that impactful and realistic milestones are achieved.

7. Conclusions and recommendations

FAO has started to modernize its IT infrastructure through the setup of a centralized Statistical Data Warehouse and FAODATA Explorer, FAO's new dissemination platform that enables access, search and sharing of structured data and metadata based on an internationally agreed SDMX standard and related tools. The project will continue its integration of FAO disseminated data, while modernizing and bringing in efficiencies in the business processes. Through this project, the quality of FAO data and statistical outputs will improve by using harmonized standards, concepts and codes toward improving data coherence, compatibility, accessibility and clarity. The organization project will take into consideration the necessary time and efforts for change in practices for a proper adoption of dissemination using SDMX.

The ongoing FAO project to leverage SDMX standard for the dissemination and exchange of data will dovetail with the ongoing efforts for greater data digitalization in Africa to catalyze innovative solutions for data production and improvements in the quality of disseminated statistics. Coordination of these interventions will be needed, before long, to maximize their cumulative impacts on making incremental increase of data quality and data user satisfaction while decreasing data management and administrative costs through countries' response burden and related statistical work.

8. QUESTIONS AND INVITATIONS TO AFCAS MEMBER COUNTRIES
AFCAS Members are invited to:
- Take note of the information reported in this paper;
- Express their views, as deemed appropriate, on the modernization project and its next steps;
- Provide comments on the Beta version of the FAODATA Explorer during the AFCAS session or by contacting the team through: data-explorer@fao.org ;
- Make recommendations on the list of FAO data assets that should be prioritized for integration in FAODATA Explorer in the next phase of the project.

9. REFERENCES

FAO, 2020. Evaluation of FAO's statistical work. Thematic Evaluation Series, 06/2020. Rome. 70 pages.
https://www.fao.org/documents/card/en?details=CA9461EN%2f

FAO, 2021. Proposal for improved governance of FAO statistical activities. Programme committee, Hundred and Thirty-second Session. 11pages.
https://www.fao.org/3/ng778e/ng778e.pdf

FAO, 2023. FAO Statistics and Data Quality Assurance Framework. Rome.
https://www.fao.org/3/cc6683en/cc6683en.pdf. [Last accessed on 23 Oct 2023].

FAO, 2023. Statistical Standard on Metadata Dissemination for FAO statistical databases, version 2.0, Rome
https://www.fao.org/3/cb9292en/cb9292en.pdf [Last accessed on 23 Oct 2023].

ISTAT, 2020. Meta-Data-Manager: The SDMX Istat Toolkit.
https://sdmx.org/?page_id=5549 [Last accessed on 19 Oct 2023].
https://github.com/SDMXISTATTOOLKIT [Last accessed on 19 Oct 2023].

SDMX, 2020. SDMX glossary version 2.1. 130 pages.
https://sdmx.org/wp-content/uploads/SDMX_Glossary_version_2_1-Final-2.docx

UNSD, 2019. Working Group on SDMX under the Inter-Agency Expert Group on SDGs.
htts://unstats.un.org/sdgs/iaeg-sdgs/sdmx-working-group/ [Last accessed on 19 Oct 2023].