| 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-Sixth Session

Libreville, Gabon, 4 – 8 November 2019

AGENDA ITEM 8

Food and Agriculture Microdata (FAM) Catalogue: A new platform for evidence based decision making

Michael Rahija
Statistician
Office of Chief Statistician

SUMMARY

The Food and Agriculture (FAM) catalogue was developed in order to disseminate micro datasets that cover topics related to FAO's mandate (i.e. food, agriculture, nutrition, etc.). The platform was launched during July, 2019 with around 350 datasets, including all the micro dataset used to compute the Food Insecurity Experience Scale (FIES) as well as datasets on food and agriculture harvested from several national platforms (Rwanda, Nigeria, and Nepal). Countries can benefit from using FAM in two main ways: by contributing their datasets to be shared via FAM, and by directly accessing data within it. This paper will briefly describe microdata and why FAM was developed, standards implemented, how countries can benefit from it, and the next steps.

I. What are microdata, and why is FAM needed?

Data collected through surveys, and administrative systems form the foundation of official statistics, and are an invaluable source for research. They are aggregated to generate national estimates by official statisticians, and analysed by researchers and policy analysts to gain scientific insights which can be translated into policy. These data commonly referred to as microdata are defined as unit-level information on individual people or entities (such as individuals, households, business enterprises, farms, or even geographic areas).

The power of microdata stems from its granularity. Because microdata contain individual level information, they allow an analyst to investigate the unique ways a certain phenomenon may affect sub-populations. For example, a particular agricultural policy may affect male and female agricultural holders differently. Likewise, a social protection scheme may benefit a particular demographic and disadvantage another. This type of analysis is impossible without highly granular datasets which allow for the analyst to stratify a dataset by a one or more variables.

FAO and member countries increasingly rely on micro datasets collected through agricultural surveys and censuses for monitoring and evaluation, tailor programming and policy interventions, conducting research, and monitoring important development trends and indicators such as small holder resilience and animal disease. Furthermore, the demand for highly granular and disaggregated microdata is exacerbated by the SDG indicators related to food insecurity (2.1.2), small productivity (2.3.1, 2.3.2), agricultural sustainability (2.4.1), risk of extinction of local breeds (2.5.2), women's ownership of agricultural land (5.a.1), food loss (12.3.1) and many others. In this regard, FAO has ongoing donor and regular program funded activities which collect farm and household survey data.

National governments and international organizations increasingly rely on microdata dissemination as a way to derive additional value. Collecting microdata is very expensive, and labor intensive. Analysing the data a single time to generate estimates or publish a report results in a significant opportunity cost. Survey data typically contain many variables, and it is infeasible that a national statistical office, or line ministry has the time and resources to fully exploit their analytical value. Accordingly, disseminating micro datasets gives others the opportunity to derive additional insights and add value.

Microdata sharing adds transparency to FAO's work. FAO derives indicators from microdata including the Food Insecurity Experience Scale (FIES), and the Global Individual Food Consumption Tool (GIFT). Providing researchers access to the microdata makes FAO's results reproducible and increases credibility. Furthermore, it can also elicit insights and methodological improvements to improve FAO's work.

Despite its importance, until this year, FAO lacked the policies and IT platform to disseminate both microdata. In this regard, with resources provided by the Bill and Melinda Gates Foundation and USAID, the Food and Agriculture Microdata (FAM) Catalogue was launched at https://microdata.fao.org during July 2019. At launch, FAM disseminated the 331 datasets used to compute the Food Insecurity Experience Scale (FIES), as well as datasets related to food and agriculture from Nepal, Nigeria, and Rwanda. To tackle the legal and administrative issues, the Office of Chief Statistician developed a Statistical Disclosure Control (SDC) Protocol, and Micro and Metadata Curation and Dissemination Protocol.

The objective of the FAM catalogue is to become the world's one-stop-shop for locating micro datasets relevant to FAO's mandate. The idea is that whenever a statistician, researcher, policy analyst, or development practitioner anywhere in the world wishes to find agricultural microdata, FAM is the first place they try. The International Household Survey Network (IHSN) serves the household survey community, while IPUMS catalogues Population and Housing Census data. Accordingly, FAM fills the agricultural microdata gap.

II. Standards

The FAM catalogue follows well-known international standards for IT, metadata, and confidentiality. Following already established standards alleviates the risk of 'reinventing the wheel', and ensures that the FAM catalogue is interoperable with other microdata dissemination platforms.

a. IT

The FAM catalogue leverages the National Data Archiving (NADA) platform that is built and maintained by the World Bank. NADA was chosen because it is the most used platform for microdata sharing worldwide which makes exchanging metadata information relatively simple between platforms. NADA was deployed as a cloud application on Amazon Web Services (AWS), and the user registration system leverages Cognito which allows external users to use Gmail accounts, or create new accounts no matter their email provider. There are many back-end features which provide IT security protection, as well as render it essentially impossible to crash.

b. Metadata

NADA leverages the Data Documentation Initiative (DDI) metadata standard. The DDI was designed to capture all of the information needed for analysing microdata. It also includes sets of controlled vocabularies. Though the list is much longer, FAM require clear and accurate information for the following minimum set of DDI elements listed below:

1. Title
2. Abstract
3. Kind of data (i.e. sample survey, census, etc.)
4. Unit of Analysis
5. Country
6. Geographic coverage
7. Sampling procedure
8. Weighting
9. Dates of data collection
10. Authority entity/primary contact
11. Access conditions

NADA further uses the Resource Development Framework (RDF) standard for cataloguing the physical datasets, and related materials (i.e .questionnaires, methodological documents, etc.). The integration of DDI and RDF is realized through a metadata document generated in XML (Extensible Mark-up Language). XML is ideal because it is a non-proprietary format which can used to generate metadata documents in a variety of more "human readable" formats (i.e. .pdf, HTML, etc.).

c. Confidentiality

The Office of Chief Statistician (OCS) developed the Statistical Disclosure Control (SDC) protocol which describes the methods and procedures for ensuring confidentiality in the datasets published through FAM. The SDC protocol cites well known tools and approaches previously published by UNECE, the World Bank, and academic literature.

The following points describe the main steps in the anonymization work flow:

1. Removal of direct identifiers and extremely sensitive variables
2. Definition of key variables, disclosure scenarios, preferred terms of access[1], and published statistics
3. Measure risk and apply disclosure limitation methods
4. Evaluate protected dataset and document
5. Approval by the Chief Statistician and release of the anonymized microdata file

As mentioned in point 2, for the time being all microdata disseminated directly through FAM will be under licensed use only. This means that users will have to fill out a short application describing the intended use of dataset prior to being granted access. OCS evaluates these requests based on two main criteria: 1. the dataset is fit for the purpose stated in the application, 2. the user comes from a credible institution which is unlikely to violate any terms of use. This policy may be reconsidered as more experience is gained.

III. How countries can use the FAM catalogue

The FAM catalogue is also intended to be used by countries in 2 main ways: 1. Increase the use and visibility of micro datasets by republishing micro and/or metadata, and 2. Provide access to micro datasets which are relevant to their work. This section will describe the benefits and procedures of these two uses.

---

[1] For the foreseeable future, all datasets will be distributed in the same way requiring users to submit an application prior to access.

a. Publishing/republishing micro and metadata

FAO's data platforms are well-known to the international community, and receive millions of hits per year. Google Analytics indicates 1,580 users, and more than 9,000 page views in the first few months since FAM's launch. Because the FAM is an aggregator for datasets all over the world, it is likely to gain much more users than any single platform increasing over time. As such, by making a datasets findable in FAM, it is likely to be exposed to a much larger audience than any single national platform. There are two main approaches to making data discoverable in FAM: 1. Sharing metadata only, 2. Sharing metadata and microdata.

Many countries already publish micro datasets in some cases even using NADA (e.g. Rwanda, and Nigeria). In these cases, a simple e-mail between OCS and the national micro dataset owner (e.g. statistical office or line ministries) is sufficient for publishing in FAM. The OCS team will either harvest the DDI metadata from the national platform, or if not in DDI, OCS will review relevant methodological documents and reports in order to develop the DDI metadata document. If the original language is not in English, FAO will translate the metadata to English. Then, the draft DDI document in pdf is always sent to the country for final approval. If the country would like the DDI XML metadata document to share through their own platform, OCS will also provide it. In this context, the <u>microdata is not shared through FAM</u>, the user is directed to the national site

Sharing both microdata and metadata requires a legal agreement. FAO's legal department is reviewing a legal template that can be signed by countries. In compliance with the legal agreement, the dataset must meet some quality standards, as well as the requirements of the Statistical Disclosure Control Protocol. Sharing both microdata and metadata may be most beneficial when the national institutional does not have the resources to set-up and/or maintain a microdata dissemination platform. In this scenario, OCS will support the country in preparing the dataset to meet quality standards, and the requirements of the Statistical Disclosure Control Protocol. Once all requirements have been met, FAO will publish the dataset under licensed use. The country may request removal of the dataset at any time.

b. Access microdata from other countries

Accessing microdata from other countries may be beneficial for several reasons. First, when implementing a questionnaire it is a common practice to look in questionnaires that have collected similar data to see how it was done. Analysing the corresponding microdata, can give insights as to how well the approach worked, identify improvements, and result in improved questionnaire design. Second, having access to neighbouring countries' microdata may prove to be useful in improving methods, and estimations. Third, by better understanding the data collection practices of similar countries, opportunities to harmonize data collection instruments, and build south-south collaboration may be identified. Lastly, countries may use microdata to inform policy design through comparative analysis of microdata from similar contexts.

Any visitor of https://microdata.fao.org can browse the data catalogue and see all of the metadata, related documents, and data dictionary without doing anything. For the datasets which are located in external catalogues, users can simply click "get microdata" and will be re-directed to the national platform where data access will be managed according to national terms of use. For the datasets located in the FAM platform, users who wish to download data must register and fill-out a data request form and data access will be granted by FAO according to the criteria mentioned earlier.

IV. <u>Next Steps</u>

Organizations that would like to contribute to FAM shall send an email to FAM-Catalogue@fao.org. Note that OCS will get in touch with a few countries which have already been identified based on the data they currently publish, or previous collaborations with FAO.

V. <u>Questions and invitations to AFCAS members</u>

AFCAS members are requested to express their views and recommendations to FAO on the following:

- What are the primary legal and political constraints for microdata dissemination, and what role can FAO play in helping to address them?

- What are the primary technical constraints (i.e. IT dissemination platform, anonymization, etc.) and what role can FAO play in helping to address them?

- For countries that already disseminate microdata, what are the main users groups and how can FAO serve their needs?