



COMMISSION DES STATISTIQUES AGRICOLES POUR L'AFRIQUE

Vingt-sixième Session

Libreville, Gabon, 4-8 novembre 2019

POINT 8 DE L'ODRE DU JOUR

CATALOGUE DES MICRO-DONNEES SUR L'ALIMENTATION ET L'AGRICULTURE : UNE NOUVELLE PLATEFORME POUR UNE PRISE DE DECISIONS AXEE SUR DES DONNEES PROBANTES

Michael Rahija
Statisticien

Bureau du Statisticien en chef de la FAO

SYNTHESE

Le catalogue sur l'alimentation et l'agriculture (FAM) a été élaboré afin de vulgariser des ensembles de micro-données sur des sujets relatifs au mandat de la FAO (alimentation, agriculture, nutrition, etc.). La plateforme a été lancée en juillet 2019 avec environ 350 ensembles de données, y compris tous les ensembles de micro-données utilisées pour calculer l'Echelle de mesure de l'insécurité alimentaire vécue (FIES) ainsi que des ensembles de données sur l'alimentation et l'agriculture provenant de plusieurs plateformes nationales (Rwanda, Nigéria et Népal). Les pays peuvent exploiter le catalogue FAM essentiellement de deux manières : en partageant leurs ensembles de données à travers le catalogue FAM et en accédant directement aux données du catalogue. Ce document décrit brièvement les micro-données et explique pourquoi le catalogue FAM a été élaboré, les normes appliquées, l'avantage que les pays peuvent en tirer et les prochaines étapes.

I. Que sont les micro-données, et pourquoi le catalogue FAM est-il nécessaire ?

Les données recueillies grâce aux enquêtes et aux systèmes administratifs constituent la base des statistiques officielles et représentent une source inestimable pour la recherche. Les statisticiens officiels les ventilent afin de générer des estimations nationales tandis que les chercheurs et les analystes des politiques les examinent pour y tirer des informations scientifiques pouvant être traduites en politiques. Ces données, souvent appelées micro-données, sont définies comme des informations au niveau de l'unité sur des personnes ou des entités (telles que des individus, des ménages, des entreprises, des exploitations agricoles ou même des zones géographiques).

L'avantage des micro-données relève du niveau de détails qu'elles peuvent fournir. Le fait que les micro-données contiennent des informations de niveau individuel, permet à un analyste d'examiner l'impact particulier qu'un phénomène peut avoir sur les sous-populations. Par exemple, l'impact d'une politique agricole spécifique sur les agriculteurs peut être différent de son impact sur les agricultrices. De même, un régime de protection sociale peut avantager un groupe démographique donné et désavantager un autre. Ce genre d'analyse est impossible sans des données très détaillées qui permettent à l'analyste de stratifier un ensemble de données suivant une ou plusieurs variables.

La FAO et les pays membres s'appuient de plus en plus sur les micro-données recueillies lors d'enquêtes et de recensements agricoles pour entreprendre des activités de suivi et évaluation, adapter la programmation et les interventions politiques, mener des recherches et suivre les tendances et les indicateurs clés de développement tels que la résilience des petits exploitants et les maladies animales. Par ailleurs, la demande de micro-données très détaillées et ventilées est accentuée par les indicateurs ODD liés à l'insécurité alimentaire (2.1.2), la productivité des petits exploitants (2.3.1, 2.3.2), la durabilité de l'agriculture (2.4.1), le risque d'extinction de races locales (2.5.2), le droit de propriété des femmes sur des terres agricoles (5.a.1), les pertes alimentaires (12.3.1) et bien d'autres. A cet effet, la FAO mène actuellement des activités financées par des bailleurs et par des programmes ordinaires, qui permettent de faire la collecte des données d'enquêtes sur les exploitations agricoles et les ménages.

Les gouvernements et les organisations internationales dépendent de plus en plus de la publication de micro-données pour en déduire une valeur supplémentaire. La collecte de micro-données est très coûteuse et nécessite beaucoup de travail. Le fait d'analyser des données une seule fois pour générer des estimations ou pour publier un rapport représente un coût d'opportunité important. Les données d'enquête fournissent généralement de nombreuses variables et il est impossible qu'un bureau national de statistique ou un ministère de tutelle ait le temps et les ressources nécessaires pour pleinement exploiter leur valeur analytique. En conséquence, la diffusion d'ensembles de micro-données permet à d'autres acteurs d'en tirer des informations et une valeur supplémentaire.

Le partage de micro-données améliore la transparence des activités de la FAO. La FAO génère des indicateurs à partir des micro-données, notamment l'Echelle de mesure de l'insécurité alimentaire vécue (FIES) et l'Outil de données sur la consommation alimentaire individuelle mondiale (GIFT). L'accès des chercheurs aux micro-données rend les résultats de la FAO reproductibles et plus crédibles. Cela peut en outre apporter des éclaircissements et entraîner l'amélioration de la méthodologie des travaux de la FAO.

Malgré son importance, la FAO n'avait pas, jusqu'à cette année, les politiques et la plateforme informatique nécessaires à la diffusion des deux micro-données. C'est pour cela qu'avec l'appui de la Fondation Bill et Melinda Gates et de l'USAID, le catalogue des micro-données sur l'alimentation et l'agriculture (FAM) a été lancé sur <https://microdata.fao.org> en juillet 2019. A son lancement, le catalogue FAM a publié les 331 ensembles de données utilisés pour calculer l'Echelle de mesure de l'insécurité alimentaire vécue (FIES), ainsi que des ensembles de données relatives à l'alimentation et à l'agriculture du Népal, du Nigéria et du Rwanda. Le Bureau du statisticien en chef a élaboré un Protocole de contrôle de la divulgation statistique (SDC) et un Protocole de conservation et de diffusion des micro-données et des métadonnées pour résoudre les problèmes juridiques et administratifs.

L'objectif du catalogue FAM est de devenir le guichet unique mondial contenant des micro-données pertinentes au mandat de la FAO. Son objectif est de faire du catalogue FAM la première référence de tout statisticien, chercheur, analyste des politiques ou praticien du développement désireux d'obtenir des micro-données agricoles. Le Réseau international d'enquêtes sur les ménages (IHSN) est au service de la communauté des enquêtes sur les ménages, tandis que l'IPUMS répertorie les données du Recensement de la population et des logements. Ainsi, le Catalogue FAM comble le manque de micro-données agricoles.

II. Normes

Le catalogue FAM est conforme aux normes internationales bien connues en matière d'informatique, de métadonnées et de confidentialité. Le respect des normes déjà établies réduit le risque d'avoir à « réinventer la roue » et garantit l'interopérabilité du catalogue FAM avec d'autres plateformes de diffusion de micro-données.

a. Informatique

Le Catalogue FAM tire parti de la plateforme NADA (National Data Archiving) créée et gérée par la Banque Mondiale. Le choix de la plateforme NADA a été influencé par le fait qu'elle est la plateforme la plus utilisée pour le partage de micro-données ; cela facilite l'échange de métadonnées entre les plateformes. La plateforme NADA a été déployée comme une application cloud sur Amazon Web Services (AWS), et le système d'enregistrement des utilisateurs utilise Cognito, qui permet aux utilisateurs externes d'utiliser des comptes Gmail ou de créer de nouveaux comptes, quel que soit leur fournisseur de messagerie. Il contient de nombreuses fonctionnalités d'arrière-plan qui assurent la sécurité informatique et empêchent le système de se planter.

b. Métadonnées

La plateforme NADA s'appuie sur la norme des métadonnées de l'initiative de documentation de données (DDI). L'initiative DDI a été conçue de sorte à prendre en compte toutes les informations nécessaires à l'analyse des micro-données. Elle comprend également des ensembles de vocabulaires contrôlés. Bien que la liste ne soit pas exhaustive, le catalogue FAM requiert des informations claires et précises sur l'ensemble minimal d'éléments DDI ci-dessous :

1. Titre
2. Synthèse
3. Type de données (enquête sur échantillon, recensement, etc.)
4. Unité d'analyse
5. Pays
6. Couverture géographique
7. Procédure d'échantillonnage
8. Pondération
9. Dates de collecte des données
10. Entité autorité/contact principal
11. Conditions d'accès

De plus, la plateforme NADA utilise la norme RDF (Resource Development Framework) pour cataloguer les ensembles de données physiques et les matériels connexes (questionnaires, documents méthodologiques, etc.). L'intégration de l'Initiative DDI et du cadre RDF s'est faite grâce à un document de métadonnées généré en XML (langage de balisage extensible). Le format XML est idéal, car c'est un format non exclusif qui permet de générer des documents de métadonnées dans une variété de formats plus « lisibles par l'homme » (par exemple, PDF, HTML, etc.).

c. Confidentialité

Le Bureau du statisticien en chef (OCS) a élaboré le Protocole de contrôle de la divulgation statistique (SDC), qui décrit les méthodes et procédures permettant d'assurer la confidentialité des ensembles de données publiés par le catalogue FAM. Le Protocole SDC fait référence à des approches et des outils bien connus publiés antérieurement par la CEE-ONU, la Banque Mondiale et des ouvrages universitaires.

Ci-dessous, les étapes clés pour rendre un flux de travail anonyme :

1. Suppression des identifiants directs et des variables extrêmement sensibles ;
2. Définition des variables clés, des scénarios de divulgation, des conditions privilégiées d'accès¹, et des statistiques publiées ;
- ↑ 3. Evaluation des risques et application des méthodes de restriction de la divulgation ;
- ↓ 4. Evaluation de l'ensemble de données et du document protégés ;
5. Approbation par le Statisticien en chef et diffusion du fichier de micro-données anonyme.

Comme indiqué au point 2, toutes les micro-données publiées directement par le catalogue FAM seront pour le moment utilisées uniquement sous licence. Cela signifie que les utilisateurs devront remplir une courte fiche de demande décrivant l'utilisation prévue pour l'ensemble des données avant de pouvoir y accéder. Le Bureau du Statisticien en chef évalue ces demandes en fonction de deux critères principaux : 1. si l'ensemble de données est adapté à l'objectif indiqué sur la demande, 2. si l'utilisateur est issu d'une institution crédible susceptible de ne pas violer les conditions d'utilisation. Cette politique peut être revue au fur et à mesure que l'on acquiert de l'expérience.

III. Utilisation du Catalogue FAM par les pays

Le catalogue FAM est également destiné à être utilisé par les pays principalement de deux manières : 1. Améliorer l'utilisation et la visibilité des ensembles de micro-données en republiant les micro-données et/ou les métadonnées, et 2. Donner accès à des ensembles de micro-données pertinents pour leur travail. Cette partie portera sur les avantages et les procédures de ces deux utilisations.

a. Publication/republishation de micro-données et de métadonnées

La communauté internationale connaît bien les plateformes de données de la FAO qui sont consultées des millions de fois par an. Google Analytics indique 1 580 utilisateurs et plus de 9 000 pages vues au cours des premiers mois suivant le lancement du catalogue FAM. Puisque le catalogue FAM est un agrégateur des ensembles de données dans le monde entier, il est probable qu'avec le temps, son nombre d'utilisateurs augmente beaucoup plus que celui de n'importe quelle autre plateforme. Par conséquent, en rendant un ensemble de données découvrable sur le Catalogue FAM, cet ensemble pourrait être exposé à un public beaucoup plus large que sur n'importe quelle unique plateforme nationale. Il existe deux approches principales pour rendre les données découvrables dans le catalogue FAM : 1. Partager uniquement des métadonnées, 2. Partager des métadonnées et des micro-données.

De nombreux pays publient déjà des micro-données, en utilisant même parfois la plateforme NADA (par exemple, le Rwanda et le Nigéria). Dans ces cas, un simple échange de courriel entre l'OCS et le propriétaire de l'ensemble de micro-données national (par exemple, un bureau de statistique ou des ministères de tutelle) suffit pour la publication des micro-données dans le catalogue FAM. L'équipe du Statisticien en chef fera soit la collecte des métadonnées DDI à partir de la plateforme nationale ou, à défaut, l'équipe examinera les documents et les rapports méthodologiques pertinents afin d'élaborer le document de métadonnées DDI. La FAO traduira les métadonnées en anglais au cas où la langue d'origine n'est pas l'anglais. Ensuite, le projet de document DDI en format PDF sera envoyé au pays d'origine pour l'approbation finale. Si le pays souhaite que le document de métadonnées XML de l'Initiative DDI soit partagé sur sa propre plateforme, l'OCS le fera également. Dans ce cas, les micro-données ne sont pas partagées sur le Catalogue FAM, l'utilisateur est renvoyé vers le site national.

Il faut un accord juridique pour partager des micro-données et des métadonnées. Le département des affaires juridiques de la FAO est en train d'examiner un modèle juridique pouvant être signé par les pays membres. Conformément à l'accord juridique, l'ensemble de données doit respecter certaines normes de qualité, ainsi que les exigences du Protocole de contrôle de la divulgation statistique. Le partage des micro-données et des métadonnées serait plus avantageux lorsque l'institution nationale ne dispose pas des ressources nécessaires pour mettre en place et/ou entretenir une plateforme de publication de micro-données. Dans ce cas, l'OCS

¹Pour un avenir prévisible, tous les ensembles de données seront répartis de manière identique, ce qui obligera les utilisateurs à soumettre une demande avant d'avoir accès aux données.

accompagnera le pays dans l'élaboration de l'ensemble de données afin de respecter les normes de qualité et les exigences du Protocole de contrôle de la divulgation statistique. La FAO publiera l'ensemble de données sous licence une fois que toutes les conditions auront été remplies. Le pays peut demander la suppression de l'ensemble de données à tout moment.

b. Accès aux micro-données d'un autre pays

L'accès à des micro-données provenant d'autres pays peut être avantageux pour plusieurs raisons. Premièrement, lors de la préparation d'un questionnaire, il est normal de consulter les questionnaires qui ont permis de collecter des données similaires pour s'en inspirer. L'analyse des micro-données correspondantes peut donner des indications sur l'efficacité de l'approche ; elle peut permettre d'identifier les améliorations et d'élaborer un meilleur questionnaire. Deuxièmement, le fait d'avoir accès aux micro-données des pays voisins pourrait s'avérer utile pour améliorer les méthodes et les estimations. Troisièmement, le fait de mieux comprendre les pratiques de collecte de données de pays similaires pourrait permettre d'identifier des possibilités d'harmoniser les instruments de collecte de données et de renforcer la collaboration Sud-Sud. Finalement, les pays peuvent utiliser des micro-données pour renseigner la conception de leurs politiques grâce à une analyse comparative de micro-données de contextes similaires.

Tous ceux qui visitent <https://microdata.fao.org> peuvent parcourir le catalogue de données pour voir toutes les métadonnées, les documents connexes et le dictionnaire de données sans faire quoi que ce soit. Pour les ensembles de données situés dans des catalogues externes, les utilisateurs peuvent simplement cliquer sur « obtenir des micro-données » pour être réorientés vers la plateforme nationale où l'accès aux données sera géré conformément aux conditions d'utilisation nationales. Pour les ensembles de données situés sur la plateforme FAM, les utilisateurs désireux de télécharger des données doivent s'inscrire et remplir un formulaire de demande de données. L'accès aux données sera accordé par la FAO conformément aux critères susmentionnés.

IV. Perspectives

Les organisations qui voudraient contribuer au catalogue FAM doivent envoyer un courrier électronique à FAM-Catalogue@fao.org. Veuillez noter que l'OCS contactera quelques pays qui ont déjà été identifiés sur la base des données qu'ils publient actuellement ou sur la base de collaborations précédentes avec la FAO.

V. Questions et invitation aux membres d'AFCAS

Les membres d'AFCAS sont invités à exprimer leurs points de vue et recommandations à la FAO sur les éléments suivants :

- Quelles sont les principales contraintes juridiques et politiques auxquelles fait face la diffusion des micro-données et quel rôle la FAO peut-elle jouer pour contribuer à les surmonter ?
- Quelles sont les principales contraintes techniques (plateforme de diffusion informatique, anonymisation, etc.) et quel rôle la FAO peut-elle jouer pour contribuer à les surmonter ?
- Pour les pays qui publient déjà des micro-données, quels sont leurs principaux groupes d'utilisateurs et comment la FAO peut-elle répondre à leurs besoins ?