

Atelier des Directeurs Généraux des Instituts nationaux de la statistique des Etats membres régionaux de la BAD sous la supervision d'AFRISTAT

sur la législation statistique et l'initiative Open data

Bamako, 24 septembre 2013

L'initiative Open Data

Freeman AMEGASHIE, septembre 2013

Introduction

Les notions de « Open data », « Big data », « Data portal » ont pris de l'essor ces dernières années dans la communauté statistique comme dans la communauté scientifique et informationnelle toute entière. Toutes ces notions tournent autour de la recherche d'une plus grande valorisation des données existantes pour répondre aux besoins d'informations des utilisateurs, des décideurs ou du grand public. Dans cette note, nous nous intéresserons surtout à l'initiative d'ouverture publique des données, Open data. Il faut aussi dire que l'ouverture des données fait partie d'un ensemble plus vaste d'initiatives intitulé « le gouvernement ouvert » visant à (i) la transparence, garant de la confiance mutuelle entre citoyens et politiques, (ii) la participation citoyenne dans la prise de décision et (iii) la collaboration pour une meilleure efficacité.

Le concept « Open data » s'entend comme l'ouverture de données aux publics. Il s'agit de rendre accessibles au public les données statistiques disponibles au sein des institutions. Ce faisant, on remplit l'une des dimensions de la qualité des données qu'est l'accessibilité. En effet, l'accessibilité de l'information statistique est définie comme la facilité avec laquelle les utilisateurs peuvent l'obtenir auprès du producteur. Cette accessibilité aux données concerne autant les indicateurs synthétiques produits par les services statistiques que les données individuelles de base collectées lors des opérations de recensements, d'enquêtes ou les fichiers administratifs.

Dans la communauté statistique, l'initiative Open Data n'est pas nouvelle. En effet, depuis quelques décennies, des actions ont été initiées pour encourager les producteurs des données pour leur large diffusion auprès du public et pour leur ouverture aux chercheurs et universitaires afin qu'elles puissent bénéficier de plus d'analyses. A cet effet, il faut noter que les producteurs de données statistiques que sont les INS et les services statistiques sectoriels n'ont pas, pour la plupart, les ressources (humaines, financières et temporelles) nécessaires pour des analyses plus approfondies des données collectées à grand frais. Mais ce faisant, les statisticiens doivent garantir la confidentialité des données individuelles collectées auprès des unités statistiques dans le cadre de recensements, d'enquêtes ou de fichiers administratifs. Cette notion de secret statistique est un pilier de la fonction statistique publique.

Conformément à son traité, l'Observatoire économique et statistique d'Afrique subsaharienne, conscient de l'importance d'une bonne utilisation des données, appuie ses Etats membres dans la création et la gestion des bases et banques de données statistiques et dans la diffusion des informations socioéconomiques relatives à ses pays. Par ailleurs, Il est important que les SSN diffusent leurs données en accord avec la législation statistique nationale. C'est pourquoi, AFRISTAT encourage ses Etats membres à élaborer et à adopter une politique de diffusion des données définissant clairement les conditions de diffusion et d'utilisation des données de façon à garantir leur confidentialité et leur représentativité.

1. Les activités de diffusion des données statistiques

Rappelons d'abord la différence que nous mettons entre les bases de données d'indicateurs statistiques et les banques de données d'enquêtes ou recensements statistiques.

Les bases de données d'indicateurs statistiques contiennent des indicateurs synthétiques calculés après un travail méthodique des statisticiens pour leur production à partir de données individuelles collectées au cours des opérations régulières ou ponctuelles. Il s'agit d'indicateurs synthétiques, appelés aussi macro-données, dont l'ensemble des métadonnées peuvent être obtenues auprès des producteurs. Les services statistiques prennent, pour la plupart des cas, des dispositions afin que ces indicateurs soient conformes aux standards et normes internationaux, de façon à garantir leur comparabilité avec les données d'autres pays. Ils sont déclinés selon des nomenclatures communes dans un souci d'harmonisation et de comparabilité.

Les banques de données regroupent les données individuelles de recensements ou d'enquêtes auprès des unités individuelles, appelées encore micro-données. Il s'agit des informations individuelles collectées directement auprès des unités primaires d'enquêtes qui n'ont pas subi d'opération de synthèse ni d'agrégation. Ce sont ces dernières, traitées et extrapolées à l'ensemble de la population étudiée, deviennent les indicateurs, objets des bases de données.

Les organisations internationales collectent et élaborent une masse considérable de données statistiques relatives aux pays pour leur utilisation propre et pour l'information du public. Souvent, ces organisations élaborent des indicateurs sur la base des estimations lorsque ces informations n'existent pas au niveau des pays. Initialement destinées aux utilisations internes, ces données n'étaient donc pas diffusées au public. Mais, depuis quelques années, ces organisations mettent ces données à la disposition de tous, en accès libre, gratuit et facile grâce aux nouvelles technologies de l'information et la communication. On pourra consulter utilement les sites de données de la Banque mondiale, <http://donnees.banquemondiale.org>, de la Division des statistiques des Nations unies (UNSD), UNData (<http://data.un.org>), de la BAD, <http://opendataforafrica.org/>.

Au niveau d'AFRISTAT, une base centralisée de données socioéconomiques des Etats membres dite AFRILDB existe depuis 2000 et permet de diffuser les principaux indicateurs statistiques des pays à un niveau agrégé ; la publication des données désagrégées est laissée à l'initiative des pays. Cette base de données a été mise en place avec le soutien de la Banque mondiale. Elle permet la diffusion des données, notamment sur le site Internet <http://www.afritat.org/publication>. En ce qui concerne les recensements et enquêtes auprès des ménages, la publication des données individuelles est du ressort exclusif des pays producteurs.

2. Collaborations et outils diversifiés

AFRISTAT, en collaboration avec des partenaires spécialisés, apporte des appuis aux pays pour organiser leurs données en bases et banques de données. Les activités développées sont relatives à la conservation des données, à la documentation, à l'insertion des métadonnées des indicateurs, à la définition de la politique de diffusion des données et à la diffusion des données. Dans le cadre de ces collaborations, plusieurs systèmes de bases / banques de données ont été mis en place dans les pays. En matière de bases de données, on peut citer 2gLDB, DevInfo, Censusingo, CountryStat, StatBase. Pour les systèmes de banques de microdonnées, les outils utilisés sont le Toolkit/Nada et IMIS/Redatam. AFRISTAT met l'accent sur le premier système.

Avec la Banque mondiale et PARIS21, AFRISTAT est membre du réseau IHSN (International Household survey network), pour la documentation et l'archivage des données grâce aux outils de **Data Management Toolkit** et la mise en place de répertoires nationaux d'enquêtes avec Nada. Grâce au Projet ADP/PARIS21, les cadres de plusieurs Etats membres ont pu bénéficier de formations à la documentation et à l'archivage des données d'enquêtes. Ce qui a abouti à la création de 8 sites nada nationaux (Burkina, Cameroun, Côte d'Ivoire, Guinée, Guinée Bissau, Mali, Niger, Sénégal). Les activités se poursuivent pour l'installation des autres sites nada.

AFRISTAT encourage et appuie ses Etats membres à évoluer vers la version de **Nada 4**. Au niveau central, il est prévu de disposer d'un portail **Nada régional AFRISTAT** avant la fin 2014.

AFRISTAT a appuyé divers partenaires dans la mise en place des systèmes et dans la formation des utilisateurs. Il s'agit de :

- **2gLDB**, de la Banque mondiale, a été installée, en plus de la Direction Générale d'AFRISTAT, dans cinq Etats membres (Cameroun, Gabon, Mauritanie, Niger, Sénégal) pour la gestion et la diffusion des indicateurs socio-économiques. Il permet une diffusion régulière de publications automatisées. Des travaux et des discussions sont en cours pour la migration vers le **Data portal** de la BAD.
- **DevInfo**, élaboré par l'UNICEF en collaboration avec le système des Nations unies afin d'aider les Nations unies et les Etats membres à suivre les progrès réalisés vers les objectifs du Millénaire pour le développement (OMD). Ce système a pour objet spécifique « de stocker les données existantes, d'identifier les lacunes dans les indicateurs OMD, de constituer un point d'entrée unique pour les données sur les indicateurs OMD, et de diffuser les informations de façon simple et attrayante. » Plusieurs Etats membres d'AFRISTAT utilisent ce support pour stocker et diffuser leurs indicateurs de développement.
- **CensusInfo**, un « système innovant et flexible de base de données pour la diffusion des résultats des recensements généraux de la population et de l'habitat développé par la Division de statistique des Nations unies, en partenariat avec l'UNICEF et le FNUAP, afin d'aider les pays à diffuser leurs résultats du recensement à tout niveau géographique pertinent, sur CD-ROM et sur le Web. »
- **CountrySTAT**, de la FAO, est un « système d'informations statistiques en ligne sur l'alimentation et l'agriculture à l'échelle nationale et infranationale. Il permet aux décideurs d'accéder aux statistiques de divers secteurs thématiques (production, prix, commerce international, consommation, etc.) en favorisant l'analyse, l'élaboration et le suivi de politiques dans le but d'éradiquer l'extrême pauvreté et la faim dans le monde ».
- **StatBase**, la base de données statistiques de la CEA, est une plate-forme Internet pour la compilation, la production et la diffusion de données sur divers indicateurs socio-économiques des pays africains. La banque de données couvre un large éventail de domaines, dont la population, l'éducation, la santé, l'agriculture, les mines, l'énergie, l'environnement, les comptes nationaux, le tourisme, le Commerce ainsi que des indicateurs des objectifs du Millénaire pour le développement (OMD).
- **IMIS-Redatam** est un système de logiciels conçu pour diffuser des fichiers de microdonnées, dans le respect du secret statistique (ou confidentialité des données). Il s'agit d'une banque de données intégrées facile d'accès (Intranet/Internet) permettant de générer des indicateurs pertinents aux niveaux national et sous-national grâce à des possibilités de calcul en ligne. Redatam a été développé et est distribué par le CELADE (le centre de démographie de l'Amérique latine et des Caraïbes de la CEPAL (Commission économique de l'Organisation des Nations-Unies pour l'Amérique latine). Redatam a bénéficié du soutien financier de plusieurs organismes, dont notamment le Centre de recherches pour le développement international (CRDI) et l'Agence canadienne de développement international (ACDI). Il est en usage par quelques Services statistiques des Etats membres d'AFRISTAT.

3. Initiative Open Data de la Banque africaine de développement

L'initiative Open Data de la Banque africaine de développement lancée en 2013 avec 20 pays pilotes a pour objectif « d'améliorer, de manière notable, la gestion et la diffusion des données en Afrique ». Le Cameroun et le Sénégal sont les deux Etats membres d'AFRISTAT figurant parmi les pays et institutions pilotes pour lesquels

la BAD développe le portail de données (Data portal). Les autres pays devraient être couverts avant la fin 2013. Courant mai 2013, une mission de la BAD a séjourné à AFRISTAT pour démarrer les travaux de la mise en place du portail de données d'AFRISTAT. Ce dernier devra ouvrir au public bientôt sa plateforme. Les deux systèmes de gestion et de diffusion de données de la BAD sont Data Portal et Open Data.

Data Portal apparaît comme une base de données d'indicateurs du pays qui permet de gérer et de diffuser facilement les données sous forme de tableaux, de graphiques ou de cartes. La manipulation des données est très aisée. Les rapports d'analyse générés peuvent être archivés.

Open Data est un entrepôt de données ouvert à tous et comportant un certain nombre d'outils spécifiques permettant de générer des indicateurs à partir de différents types de données (démographie, commerce extérieur, comptes nationaux, enquêtes, etc.).

Selon la BAD, « Open Data est un outil qui permet d'extraire les données de la base, de créer et de partager ses propres rapports personnalisés, ainsi que de visualiser les informations liées à des thématiques, à des secteurs ou à des pays, sous forme de tableaux, de graphiques et de cartes. »

La plate-forme facilite également la collecte, l'analyse et la mutualisation des données entre les pays, la BAD et les autres partenaires internationaux au développement. Elle offre aux utilisateurs la possibilité de faire leurs propres requêtes stockables et réutilisables. Cette initiative ne concerne que les indicateurs statistiques.

4. Initiative Open Data de la Banque mondiale

« L'engagement de la Banque envers une plus grande ouverture émane également de sa volonté d'accroître l'appropriation par le public, la participation d'un vaste éventail de parties prenantes et la conclusion de partenariats dans le domaine du développement. En tant qu'institution du savoir, la première mesure de la Banque mondiale est de partager ses connaissances gratuitement et ouvertement. »

C'est pourquoi depuis quelques années, la Banque a été une des premières institutions internationales à mettre en accès libre et gratuitement un ensemble de données de différents domaines du développement à l'échelle mondiale.

Le Catalogue de données de la Banque mondiale permet d'accéder à plus de 2 000 indicateurs provenant de sources diversifiées, voir <http://donnees.banquemondiale.org>.

5. La position d'AFRISTAT en matière de diffusion des données

AFRISTAT adhère parfaitement à ces initiatives d'ouverture des données au public qui donnent plus de visibilité au travail des services statistiques. Elles permettent également d'enrichir le débat sur la situation économique et sociale des pays par une plus grande utilisation des données. Toutefois, s'agissant de statistiques officielles sur lesquelles se basent les décideurs pour justifier les politiques, il est important qu'un encadrement soit fait des utilisations des données et de leur origine. Ci-après, les préconisations d'AFRISTAT selon le type de données à diffuser :

5.1. Cas des indicateurs / macro-données / bases de données :

- AFRISTAT milite pour la publication des indicateurs officiels produits par les SSN ;
- Les SSN sont les seules institutions qui doivent publier l'information officielle. Les autres diffuseurs ne doivent servir que de relais ;
- AFRISTAT publie sur son site ou (dans ses revues) des données officielles agrégées et laisse les SSN publier les données détaillées.

5.2. Cas des micro-données / Banques de données :

- AFRISTAT milite pour une plus grande valorisation des micro-données par des exploitations plus poussées et la production d'analyses par d'autres spécialistes, chercheurs et universitaires, opérateurs économiques et sociaux. Pour cela, il œuvre pour donner aux SSN les capacités de documentation et d'archivage des données et pour la diffusion des données. L'outil privilégié actuellement est le Data management Toolkit du réseau IHSN couplé avec le système Nada ;
- Le SSN sont les seules institutions qui doivent donner l'autorisation à l'utilisateur qui en fait la demande. Pour cela, le SSN doit adopter une politique de diffusion définissant notamment les différents types d'accès aux données et aux métadonnées.
- Différents types d'accès sont possibles :
 - Accès limité aux métadonnées : l'utilisateur ne peut pas accéder aux données. Mais, il a accès à toute la documentation possible ;
 - Accès limité aux données :
 - ✓ Extraction des données : le producteur ne propose que des fichiers de données extraites, un échantillon. Il ne fournit pas l'ensemble des données.
 - ✓ Exploitation sur place : le producteur met les données à la disposition des utilisateurs dans un environnement sécurisé où ces derniers réalisent des analyses sans possibilité d'emporter les données ;
 - Accès total sans limitation : l'utilisateur peut accéder à l'ensemble des données et métadonnées.
- Dans tous ces cas, l'Institution responsable des données prend toutes les mesures nécessaires garantissant la confidentialité des données individuelles et une bonne utilisation de ces données. A cet effet, il devra d'abord s'assurer de l'anonymisation des données. Ensuite, il devra conclure des protocoles d'accord pour l'utilisation des données à signer avec les utilisateurs. Toutes les analyses réalisées à partir des données du SSN doivent recevoir l'autorisation préalable du CNS, par l'intermédiaire de l'INS, qui en est le secrétariat technique, avant leur publication.

AFRISTAT, en tant qu'Observatoire économique et statistique, par son traité, doit recevoir les fichiers des pays à double titre. Il a d'abord à exercer la fonction d'archivage pour garantir les pays contre toute perte de données. Le traité instituant AFRISTAT et signé par les Etats membres, oblige ces derniers à lui fournir les bases et banques de données officielles. Ensuite, en tant qu'observatoire, il peut réaliser des études et analyses sur ces données pour ses travaux internes ou au bénéfice des Etats. Dans ce dernier cas, AFRISTAT devra présenter les résultats des travaux réalisés sur la base de ces données aux pays propriétaires et obtenir leur accord avant toute publication.

Un autre avantage pour les pays, en filigrane, est la réduction du temps consacré à répondre aux questionnaires réguliers des institutions internationales, par la possibilité que ces dernières auront maintenant à collecter directement les informations sur le portail de données de l'institution.

Conclusion et perspectives

AFRISTAT est disposé à collaborer avec les autres partenaires internationaux au développement pour appuyer ses Etats membres dans les activités menant à une ouverture publique des données statistiques qu'ils produisent en les préservant des inconvénients afin qu'ils tirent plutôt des avantages de l'initiative Open Data.

Il devra déployer des actions pour aider ses Etats membres à adopter une politique de diffusion des données en harmonie avec les lois statistiques nationales.