

Les métiers de la data au service de la statistique publique africaine

Logiciels propriétaires coûteux, cycles de publication trop longs, faible reproductibilité : comment les métiers de la data et le package open source statAfrikR modernisent les instituts nationaux de statistique (INS) d'Afrique.

Par Dikers AMOKO · 8 juillet 2026 · statafrikr.org

EN BREF · IN BRIEF

FR — La numérisation transforme la production statistique mondiale, et les métiers de la data en deviennent des acteurs incontournables. Les INS africains doivent intégrer ces compétences pour rester pertinents. statAfrikR — package R open source publié sur le CRAN — apporte une réponse souveraine et pérenne : un outil gratuit, fonctionnel hors ligne, couvrant tout le cycle de production statistique.

EN — Digitisation is reshaping how official statistics are produced, and data professions are now central to the value chain. statAfrikR — an open-source R package on CRAN — offers a sovereign, sustainable answer: a free, offline-capable tool covering the whole statistical production cycle.

La statistique publique africaine se trouve à un moment charnière. La révolution numérique bouleverse les méthodes de production, d'analyse et de diffusion des données, tandis que les INS africains affrontent des défis structurels profonds. La question n'est plus *si* la modernisation aura lieu, mais *avec quels outils* — et au service de qui.

Open source GPL-3	Publié sur le CRAN	Hors ligne sans connexion	5 modules de bout en bout	Multi-pays adapté aux INS	Reproductible Git & tests
-----------------------------	------------------------------	-------------------------------------	-------------------------------------	-------------------------------------	-------------------------------------

0 1

La statistique africaine à un carrefour

Des licences qui coûtent cher

Les logiciels propriétaires imposent des licences onéreuses et un code fermé, impossible à adapter aux réalités locales — une charge lourde pour des budgets contraints.

Des cycles trop longs

Le délai entre la fin de collecte et la publication reste élevé : traitements manuels, faible automatisation, peu de programmation avancée.

Une reproductibilité fragile

Traitements interactifs, code non documenté, ni contrôle de version ni tests : vérifier, mettre à jour et transmettre devient très difficile.

0 2

Le halo des métiers de la data

Ces profils ne remplacent pas le statisticien : ils enrichissent ses capacités. On les situe le long d'un axe allant des techniques informatiques aux usages métier (Galiana & Lefebvre, 2026).

Data engineer

Intégration des données (pipelines), déploiement, MLOps. Orienté infrastructure : rend un traitement robuste et jouable.

Data scientist

Analyse, entraînement des modèles, préparation et labellisation. Interface technique/métier — outillé en priorité par statAfrikR.

Data analyst

Exploration, visualisation, recommandations. Orienté décision : relie les indicateurs à l'action publique.

Pour les INS africains, ces rôles peuvent être exercés par les statisticiens eux-mêmes, progressivement formés à R, ou par de nouvelles recrues spécialisées.

03 R, standard mondial — et statAfrikR

R est le langage de référence de la statistique mondiale : utilisé par l'INSEE, Eurostat, l'ONU, la Banque Mondiale et le FMI, il est porté par une communauté internationale et un très riche écosystème de packages. **statAfrikR** en est l'adaptation africaine la plus avancée : publié sur le **CRAN**, il réunit plusieurs modules complémentaires et est déjà adopté par une communauté grandissante de statisticiens africains. Il fonctionne **hors ligne**, en français, sur des ordinateurs modestes.

« Un INS dont les données sont propres, traçables et harmonisées est un INS qui inspire confiance. Les décisions de politique publique ne sont jamais meilleures que les données sur lesquelles elles reposent. »

Quatre valeurs guident le projet : **souveraineté statistique**, **reproductibilité**, **harmonisation** (AFRISTAT, ONU, ODD, Agenda 2063) et **inclusion** (gratuité GPL-3, hors ligne, matériel modeste).

04 Un package qui couvre tout le cycle

De l'import des données brutes au rapport final, cinq modules s'enchaînent comme un pipeline reproductible.



Un cycle reproductible : chaque étape est rejouable et documentée.

Module	Rôle	Fonctions clés
Collecte	Importer et normaliser des sources hétérogènes	<code>import_kobo()</code> , <code>import_odk()</code> , <code>import_stata()</code> , <code>import_cspro()</code>
Traitement	Contrôler la qualité et préparer les données	<code>check_na()</code> , <code>nettoyer_libelles()</code> , <code>recoder_variable()</code> , <code>valider_qualite_donnees()</code>
Analyse	Produire statistiques et indicateurs	<code>appliquer_ponderations()</code> , <code>stat_descr()</code> , <code>calcul_ipm()</code> , <code>calcul_idh()</code>
Visualisation	Cartographier et illustrer	<code>theme_ins()</code> , <code>palette_ins()</code> , cartes thématiques PNG/SVG

Diffusion

Générer rapports et métadonnées

```
generer_rapport(),  
generer_metadonnees_ddi(),  
exporter_sdmx(), anonymiser_donnees()
```

Et la suite — v0.2.0. De nouveaux indicateurs (pauvreté FGT, ODD, pauvreté multidimensionnelle), une cartographie hors ligne étendue, un module comptabilité nationale et un tableau de bord interactif.

0 5 Des cas d'usage concrets

EHCVM — Burkina Faso

Automatisation des analyses de pauvreté, des indicateurs ODD, des tableaux institutionnels et des cartes statistiques à partir des données d'enquêtes auprès des ménages.

RGPH-4 — Tchad

Automatisation de la production d'indicateurs démographiques, de tableaux normalisés, de cartes thématiques et de rapports institutionnels à partir des données du recensement national.

RGPH — RCA

Automatisation des indicateurs démographiques, des tableaux normalisés, des cartes thématiques et des rapports institutionnels issus du recensement général de la population.

Votre INS conduit une EHCVM, un RGPH ou un suivi des ODD ? statAfrikR peut automatiser votre chaîne d'analyse et de production statistique — des données brutes aux rapports et cartes. Écrivez-nous via statafrikr.org.

06 La codification automatique par machine learning

L'INSEE a industrialisé un modèle qui attribue automatiquement le code d'activité d'une entreprise à partir du libellé rédigé en langage naturel. L'intérêt n'est pas seulement le gain de temps : c'est la **démarche** — expérimenter, mettre en production de façon fluide, puis surveiller la performance dans la durée (approche MLOps).

Transposée à l'Afrique : à partir de descriptions libres saisies dans un RGPH ou une EHCVM, un modèle pourrait attribuer le **code d'activité** (CITI/ISIC) et le **code de profession** (CITP/ISCO). Le point clé est la **surveillance du modèle** : entraîné sur des données passées, il perd en pertinence quand la réalité change (économie des plateformes, livraison à domicile, libellés inédits). Un cycle d'amélioration continue — collecte de cas annotés, réentraînement — devient indispensable. C'est le type de chantier expérimental que statAfrikR inscrit à sa feuille de route.

07 Reproductibilité : Git, tests, CRAN

Le **versionnement du code** est l'une des transformations les plus profondes qu'apporte la culture data. Avec **Git**, chaque modification est tracée, la collaboration devient fluide et les résultats reproductibles : le code cesse d'être un fichier isolé pour devenir un **bien commun institutionnel**. Chaque fonction de statAfrikR est couverte par des **tests automatisés**, et la publication sur le **CRAN** soumet l'outil aux mêmes exigences que les packages de l'INSEE, d'Eurostat ou de la Banque Mondiale — avec la transparence de l'open source.

08 Cap 2026-2030

Les prochaines versions apportent les indicateurs les plus demandés (pauvreté monétaire et multidimensionnelle, ODD), une cartographie hors ligne étendue, un module de comptabilité nationale et un tableau de bord interactif. Autour du package : la **statAfrikR Academy** (formation certifiante) et une **communauté panafricaine** sur Discord, premier réseau de statisticiens R du continent.

Une modernisation souveraine

La transformation ne se décrète pas : elle se construit, progressivement, avec des outils adaptés, des agents formés et une vision institutionnelle claire. statAfrikR n'est pas un outil parmi d'autres — c'est une infrastructure statistique pensée pour les réalités africaines, développée par des statisticiens africains, maintenue par une communauté africaine et publiée sous licence libre.

Sources & références. Galiana, L. & Lefebvre, O. (2026). *Qu'apportent les métiers de la data à la statistique publique ?* Courrier des statistiques, n°15, INSEE. — Foster, Greer & Thorbecke (1984), *Econometrica*. — Alkire & Foster (2011), *Journal of Public Economics*. — Amoko-Mongamo, D.-D. (2026). *statAfrikR*, CRAN.