

Blog

Échos sur la gouvernance des données

Renforcer le partage des données pour une agriculture intelligente prospère





L'agriculture intelligente consiste à utiliser un flux d'information et de l'intelligence artificielle pour alimenter la prise de décision en vue de hausser la productivité et mieux gérer les contraintes de toute la chaîne de valeur. Malheureusement, elle peine à être une réalité d'échelle dans le contexte africain faute de données. Pourtant, il semble que des initiatives de partage de données existent mais les informations partagées seraient non-adaptées au contexte africain, insuffisantes, incomplètes ou encore méconnues du grand public. Par conséquent, les praticiens de l'IA rencontreraient souvent une série d'obstacles pour l'acquisition et la valorisation de ces types de données.

Cette cartographie de la situation, simple au premier coup d'œil, n'est pas si triviale à faire évoluer car elle résulte de la combinaison de divers facteurs qu'il convient de bien maîtriser. Parmi ces facteurs nous énumérons dans le cadre de ce blog la culture de la donnée, la théorie du changement, la maîtrise des besoins en données et enfin l'identification des bonnes sources de données.

Culture de la donnée

La culture digitale doit être installée au sein de toute une population pour espérer une transformation numérique réussie et à des échelles suffisantes au profit du développement économique d'un pays. Avec les technologies émergentes, l'usage du numérique donne de plus en plus place à une agriculture intelligente dont l'essor dépend, en bonne partie, sur la maîtrise de la donnée. Il sied alors de penser à développer une culture de la donnée pour entreprendre la révolution de l'agriculture intelligente dans des pays comme ceux de l'Afrique. Pour ce faire, l'identification des savoirs et savoir-faire sur la donnée devient un enjeu majeur pour associer un contenu à cette culture et élaborer une ligne de conduite pour son appropriation par les parties intéressées. Cette ligne de conduite devrait combiner des campagnes de

sensibilisation dans plusieurs langues, une intégration dans les curricula de modules de formation, une culture de la documentation et une gouvernance claire des données dans les organisations, des mécanismes incitatifs de partage des informations eu égard des cadres réglementaires et normatifs, un dispositif de création de champions en données, etc.

Théorie du changement

Aucune transformation ne peut prétendre à un grand impact et encore moins prospérer dans le temps sans une conduite de changement bien élaboré et minutieusement implémentée.

C'est pourquoi, qu'au-delà d'avoir une stratégie sur la donnée ou des politiques incitatives de la libéralisation des données, il est important de sensibiliser sur l'importance des données et de leur apport à l'agriculture intelligente.

Dans cette dynamique, le postulat de base que l'on pourrait poser est qu'une collecte régulière d'informations, faite avec et pour les agriculteurs, contribue à un environnement favorable pour la valorisation des données. Une première étape serait d'élaborer un cadre d'utilisation des données (CUD) compréhensible et accessible à toute partie intéressée. Une deuxième étape consisterait à implémenter une palette d'éléments catalyseurs (PEC), lesquels sont considérés comme des intrants disponibles localement à savoir des espaces agricoles pourvoyeurs d'informations, un capital humain capable de collecter les informations et/ou d'appliquer les décisions et les moyens technologiques connexes.

En résumé, la théorie proposée fait de l'agriculteur le point de départ comme celui d'arrivée de la chaîne décisionnelle sur la donnée. Donc, il faut s'assurer que l'agriculteur ait accès à des outils/interfaces lui permettant d'initier ou de par-

ticiper à la collecte comme également de comprendre la décision déduite de l'analyse des informations dont il a contribué à la collection. Cette approche fait appel à la culture de la donnée décrite précédemment et qui est tributaire d'une bonne culture digitale.

Maitrise des besoins en données pour l'IA

Le problème le plus crucial après avoir parlé des deux facteurs précédents est de définir les attentes des praticiens de l'IA. En effet, c'est sur la base de ces attentes que la collecte devrait être spécifiée pour disposer des données en quantité et en qualité en vue de répondre à une ou des problématiques spécifiques de l'agriculture. Pour faciliter la maitrise des besoins des praticiens, une revue documentaire et des pratiques des spécialistes de la donnée nous a permis d'isoler huit critères ou dimensions qui pourraient contribuer à mieux satisfaire les attentes. En d'autres termes, on peut utiliser ces dimensions pour définir un protocole de collecte ou pour classifier une source de donnée. La figure 1 illustre l'ensemble ces critères à l'image d'une araignée.

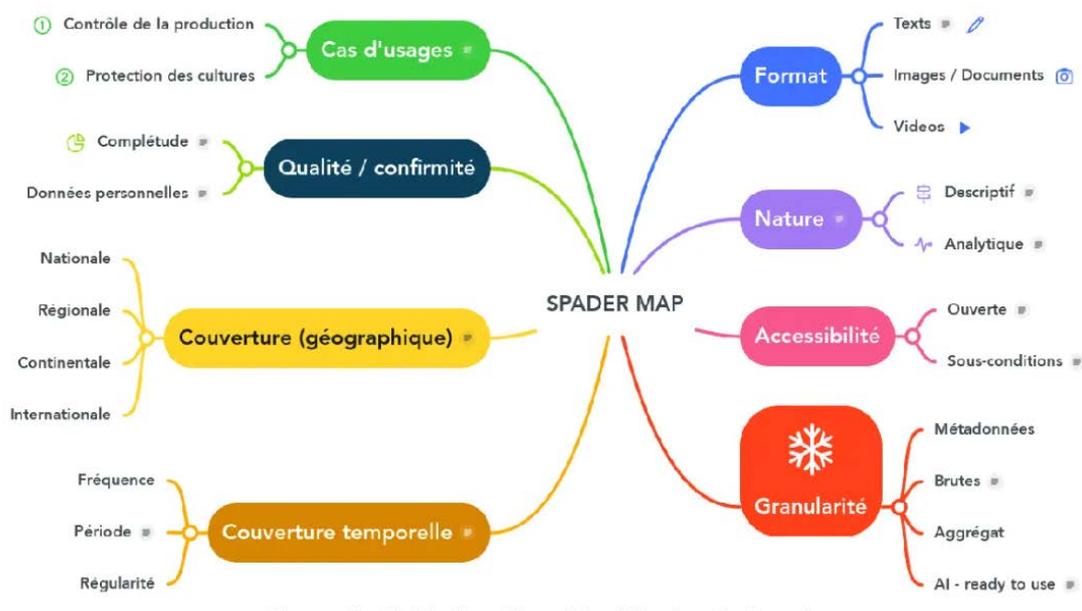


Figure 1 : Guide de collecte/classification de données



Les dimensions sont souvent une combinaison d'un ensemble de sous-dimensions qu'il convient d'expliquer ci-après.

Format : Cette dimension désigne le format des données qui peut être sous forme de texte, d'images ou de vidéos. Les formats de type texte sont entre autres les fichiers avec les extensions .csv, .xlsx, .json, .xml, etc ;

Nature des données : Elle concerne le fait que les données peuvent être à visée descriptives ou analytiques ;

Accessibilité : Elle fait référence à l'ouverture des données et/ou les conditions d'accès qui peuvent se faire via une contractualisation ou le respect d'un certain nombre de critères ;

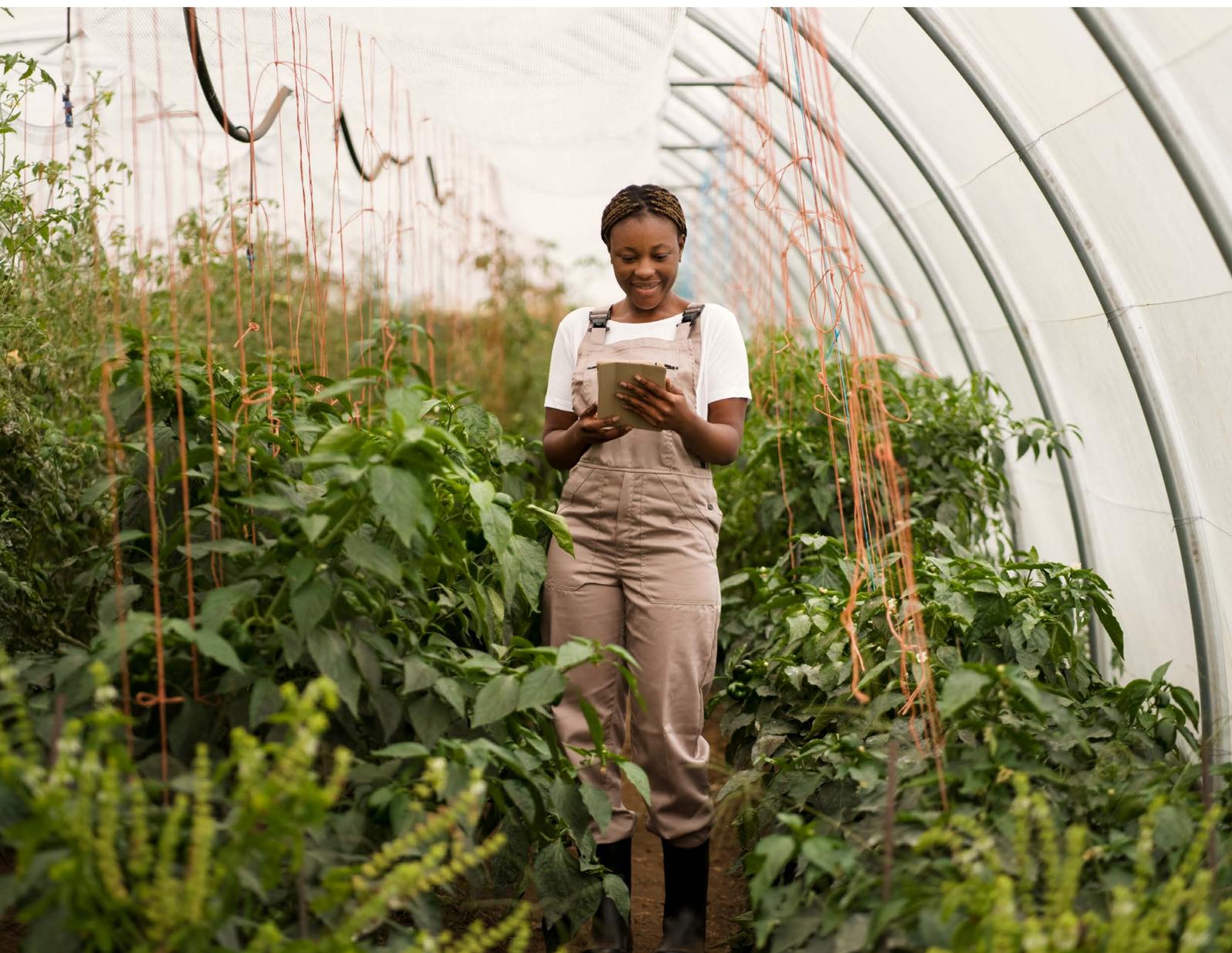
Granularité : Elle indique si les données sont de type métadonnées, données brutes, agrégées, ou dans le format directement utilisable par des algorithmes IA (ce que nous désignons par AI-ready to use) ;

Qualité/conformité : La qualité est mesurée par la complétude et le caractère officiel de la source. La complétude est assurée si les données sont bien renseignées et disponibles pour chaque variable ou dimension. L'officialité renvoie au caractère officiel et/ou non-officiel de la source et la conformité au respect de la réglementation sénégalaise sur les données à caractère personnel et/ou à la loi sur la statistique officielle.

Couverture géographique : Elle indique si les données sont relatives à une zone géographique spécifique. Pour notre étude, nous avons considéré quatre types de zones : nationale, régionale, continentale ou internationale ;

Couverture temporelle : Elle renseigne sur la périodicité, la fréquence et la régularité de la production des données en termes de structuration ;

Cas d'usages : Il renvoie aux types d'usages considérés dans le cadre de cette étude. Concrètement, il s'agit des données produites sur la production et/ou la protection des cultures.



Identification des plateformes de données

Pour identifier les bonnes sources, la dimension relatif au cas d'usage devrait être privilégié puisqu'une source n'a de valeur ajoutée que lorsqu'elle contient des données permettant de résoudre un problème. Pour le cas de cet article, il s'agirait des informations permettant d'utiliser l'IA pour améliorer la production et la protection des cultures.

Dans cette perspective, nous avons défini avec l'aide des experts en agriculture un cadre d'analyse faisant intervenir les interactions entre la plante, l'environnement et les moyens de production. Cette relation est définie sur la base que la production et la protection des cultures peuvent être améliorées en s'appuyant sur le type de plante, l'environnement et les moyens de production à disposition. Ainsi, nous avons identifié un ensemble de variables et sous-variables comme illustré par le schéma suivant :

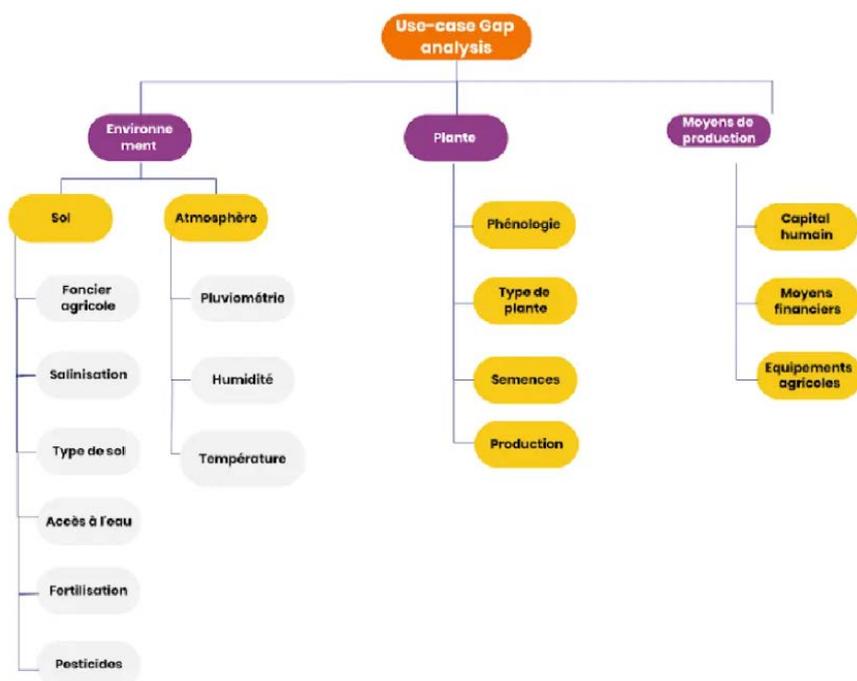


Figure 2 : Guide de collecte de données pour l'agriculture intelligente

• **L'environnement** est décrit avec deux variables à savoir le sol et l'atmosphère. Le sol est décrit avec six sous-variables pour apprécier la qualité de ce dernier et/ou avoir une idée de son état optimal nécessaire à une bonne production. L'atmosphère est décrite sur la base des sous-variables température, humidité et pluviométrie.

• **La plante** est caractérisée par les sous-variables phénologie, semences (certifiées, améliorées ou non-certifiées), production et spéculation.

L'environnement est décrit avec deux variables à savoir le sol et l'atmosphère. Le sol est décrit avec six sous-variables pour apprécier la qualité de ce dernier et/ou avoir une idée de son état optimal nécessaire à une bonne production. L'atmosphère est décrite sur la base des sous-variables température, humidité et pluviométrie.

La plante est caractérisée par les sous-variables phénologie, semences (certifiées, améliorées ou non-certifiées), production et spéculation.

1. Situation 1 : Il existe des données qui couvrent les besoins de l'IA dans le domaine agricole.

Quels mécanismes de vulgarisation et de gouvernance faut-il développer pour rendre publiques et accessibles ces données ?

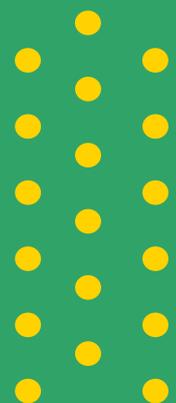
Sous quel format, faudrait-il partager ces données pour les rendre utilisables par l'IA ?

Quelles politiques de sécurité faudrait-il élaborer pour accompagner l'exploitation intensive et démocratisée de ces données sans compromettre les droits liés au numérique ?

2. Situation 2 : Il n'existe pas de données suffisantes dédiées pour l'usage de l'IA dans le domaine agricole.

Quels besoins en collecte de données faudrait-il couvrir pour développer l'agriculture intelligente en Afrique subsaharienne ?

Quelle stratégie de gouvernance des données faudrait-il déployer pour des usages efficaces, responsables et sécurisés ?



CONTACTS:

Idrissa SARR

E-mail: idrissa.sarr@ucad.edu.sn
Centre universitaire de Recherche et
de Formation aux Technologies de l'Internet.
Université Cheikh Anta Diop de Dakar

www.curi.sn

Niyel

E mail: info@niyel.net

www.niyel.net

