



COMMISSION 5

Gestion des systèmes d'information financière

Digitalisation et implication

Installation d'un datalab à la Direction Générale des Douanes de Côte d'Ivoire

Arouna N'Dia, Sous-Directeur des Etudes Economiques de la Direction des Statistiques et des Etudes Economiques

Clément Dupont, Chef du projet DATAFID, Expertise France

Version édition

1. Conception et contenu

1.1. Présentation du projet et diagnostic de démarrage

Le projet DATAFID (DATA dans les Administrations Fiscales et Douanières), financé par la Direction Générale du Trésor français, et mis en œuvre par Expertise France dans plusieurs pays d'Afrique de l'Ouest (Côte d'Ivoire, Mauritanie, Niger et Togo) est une initiative pilote, exploratoire, qui vise à répondre à la question suivante : comment accompagner la valorisation des données dont disposent les administrations fiscales de la sous-région pour qu'elles améliorent la mobilisation des ressources intérieures ?

Afin d'identifier la meilleure approche, le projet a d'abord réalisé un état des lieux de l'utilisation des données dans les administrations concernées. La méthode a combiné des visites terrain pour rencontrer les parties prenantes et un travail à distance pour procéder à des études documentaires et analyser des réponses à un questionnaire administré en ligne sur l'état de la digitalisation des entreprises. Ce questionnaire a repris les métriques contenues notamment dans le modèle de maturité de l'OCDE, qui évalue le degré d'utilisation de fonctions analytiques par les administrations fiscales¹.

¹ <https://www.oecd.org/tax/forum-on-tax-administration/publications-and-products/analytics-maturity-model.pdf>



Réformes des finances publiques en Afrique francophone

Le questionnaire s'intéressait en particulier aux ressources, usages et compétences en matière de politique de la donnée.

Les analyses aux questionnaires ont permis d'établir plusieurs constats :

- Les compétences internes des administrations sont concentrées sur le développement et l'exploitation. A l'inverse, les fonctions d'analyse ou d'ingénierie de données sont peu présentes (voir Figure 1 : pourcentage d'administrations disposant de compétences en interne) ;
- Les architectures d'analyse des données sont relativement basiques, dans le sens où elles sont principalement constituées de bases de données. Peu d'administrations disposent de lac ou entrepôt de données (voir Figure 2 : type d'architecture en place pour l'exploitation des données).

Figure 1 : pourcentage d'administrations disposant de compétences en interne

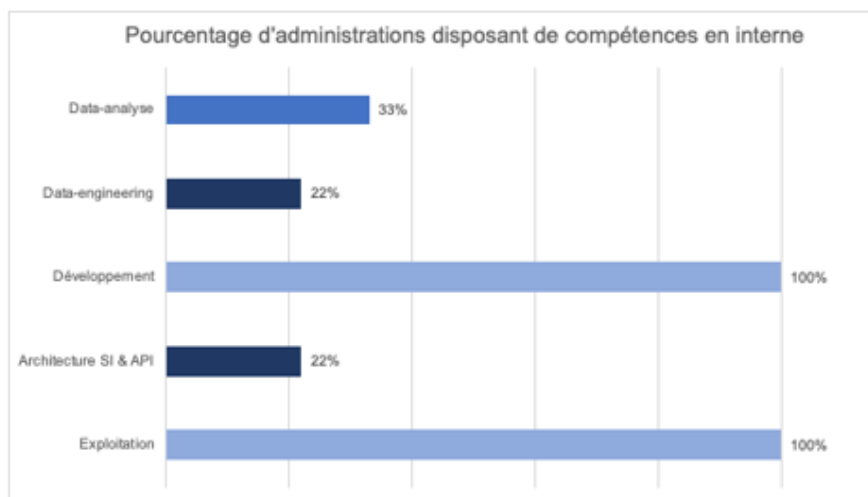
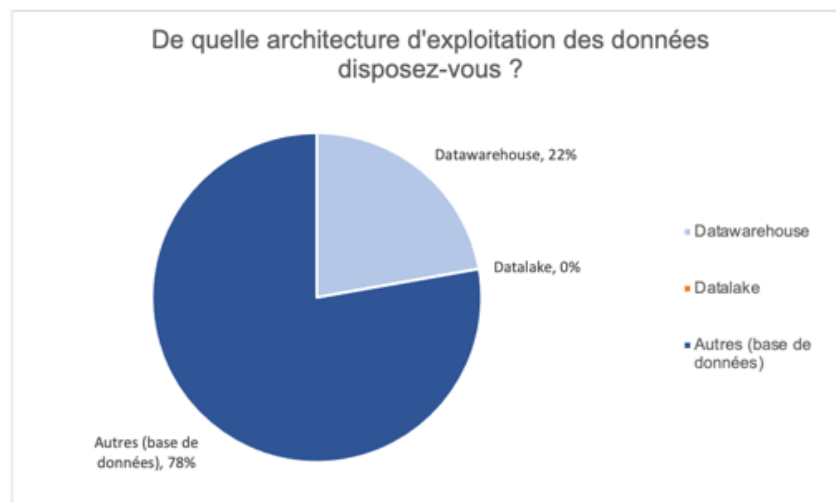


Figure 2 : type d'architecture en place pour l'exploitation des données



Par ailleurs, au cours des entretiens avec les parties prenantes, les experts ont observé des efforts relativement soutenus de digitalisation des procédures fiscales et douanières, générant en retour des volumes de données importants. Les discussions ont toutefois conclu à une sous-utilisation chronique de ces mêmes données.

1.2. Construire des cas d'usage pour atteindre les objectifs

La phase de démarrage ayant permis de mieux saisir les dynamiques en place, les experts ont décidé d'opter pour une approche de développement de cas d'usage de la science des données. Un cas d'usage peut être défini comme suit : il s'agit d'identifier un domaine, une tâche, une activité problématique et dont la résolution passe par une meilleure utilisation de la donnée. Dans notre cas, l'approche par les cas d'usage procurait plusieurs avantages :

- Il s'agit d'une approche technique, par le bas, qui permettait rapidement de rentrer dans des considérations opérationnelles, au plus près des producteurs et utilisateurs de la donnée, pour identifier et comprendre les problèmes rencontrés ;
- La construction d'un cas d'usage comprend un fort volet plaidoyer car elle permet de faire rapidement prendre conscience de la valeur de la donnée, en proposant des « quick wins », c'est-à-dire des résultats rapides ;

Réformes des finances publiques en Afrique francophone

- C'est une approche « pratique » qui a l'avantage de mobiliser davantage qu'une approche privilégiant le renforcement de capacités.

En juin 2022, le projet a donc organisé un atelier d'identification de cas d'usage en présence d'une vingtaine de partenaires du projet, issus des administrations fiscales et douanières et occupant diverses fonctions (informaticiens, statisticiens, inspecteurs des douanes et des impôts) pour permettre une approche aussi large que possible. Les organisateurs comprenaient eux aussi différents profils : conseillers en transition digitale, représentant de l'administration fiscale française pour partager une expérience « avancée », représentant de l'administration fiscale sénégalaise pour partager une expérience « voisine », représentant de l'Organisation Mondiale des Douanes, juriste spécialisé dans le numérique. Les méthodes et techniques de l'atelier ont alterné entre des présentations de cas d'usage en place dans des contextes similaires, et des ateliers participatifs pour permettre l'identification de situations problématiques, et leurs réponses potentielles. Les travaux ont abouti à l'identification de grandes familles de cas d'usage, gravitant principalement autour du ciblage des contribuables, de l'échange de données entre administrations, et de la prévisibilité des recettes.

Au terme de cet atelier, les experts ont débuté une série de visites dans les pays concernés pour affiner les cas précis et les méthodes de construction afin de garantir la pleine implication des partenaires. Ils se sont basés sur les documentations nationales (plan stratégique, études précédentes) et sur des échanges avec le management des administrations. En parallèle, le projet a commencé la co-construction de modules de formations en ligne sur l'utilisation de la science des données au profit des administrations fiscales et douanières, pour assurer la montée en compétence des partenaires.

1.3. L'outil clé pour atteindre les objectifs du projet : le datalab

Assez rapidement toutefois, le projet s'est retrouvé face à une question déterminante : dans un contexte de faible degré de maturité, quels outils mobiliser pour faire prendre conscience du potentiel de la donnée ? Dans l'ensemble, les pays partenaires manquaient d'outils et d'infrastructures pour stocker les données, s'entraîner à leur



Réformes des finances publiques en Afrique francophone

manipulation, et les faire circuler. Fallait-il se procurer des serveurs et les installer directement au sein des départements SI des administrations ? Comment gérer les enjeux de souveraineté et de sécurité liés aux données et informations? Comment dimensionner les besoins en puissance de calcul, comment être sûrs que les infrastructures/outils seront vraiment utilisés ? Comment faciliter la documentation technique des projets réalisés ?

En confrontant ces questions aux contraintes du projet, les équipes ont conclu qu'il n'était pas opportun de s'engager dans la fourniture de machines (serveurs, machines virtuelles) pour chaque pays tant que les besoins en manipulation de données n'étaient pas suffisamment dimensionnés. L'alternative : préférer une approche simple, pratique, permettant aux équipes de s'exercer à des techniques nouvelles et d'explorer le potentiel de leurs données pour montrer des résultats avant de décider d'un système plus industriel.

C'est à ce moment qu'est apparue l'idée d'utiliser un datalab, c'est-à-dire une plateforme qui offre un environnement de travail autonome pour des datascientists et statisticiens. L'équipe s'est projetée sur Onyxia, solution open source, et donc gratuite, construite par l'INSEE (Institut National des Statistiques et des Etudes Economiques de France) et mise à disposition de professionnels de la data en France et dans le monde². Cette plateforme comprend un catalogue de services de datascience, la possibilité de stocker des données de manière flexible et l'accès à une communauté.

Cette plateforme étant open source, le code et les éléments développés sont réutilisables et adaptables par tous. Les experts DATAFID ont décidé d'adapter Onyxia à leurs besoins, en prenant en considération plusieurs principes :

- Commencer par une instance mutualisée à l'ensemble des pays partenaires afin de tester les usages et identifier ensuite les besoins spécifiques en traitement de données ;
- Prendre en compte les enjeux de sécurisation des données et éviter la fuite des données sources ;

² <https://www.onyxia.sh/>

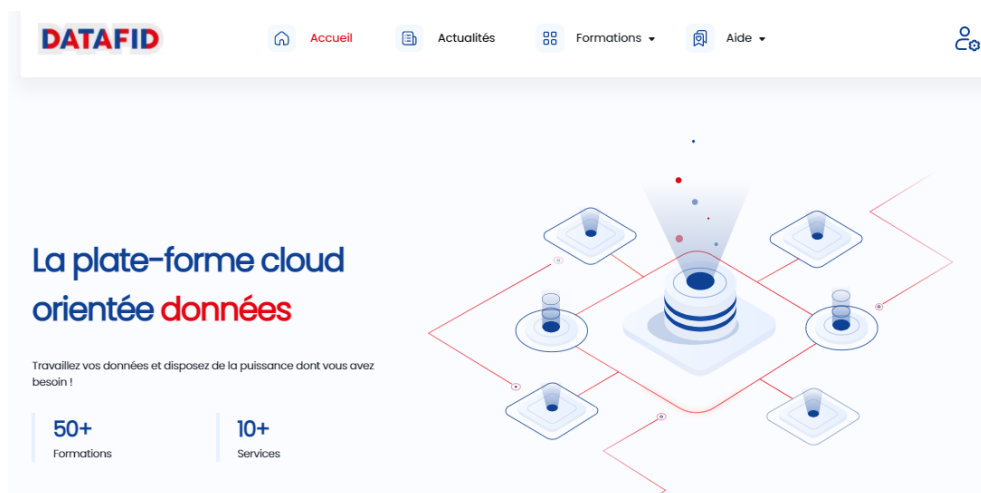


Réformes des finances publiques en Afrique francophone

- Favoriser l'acquisition de compétences et le partage de connaissances.

La première version du datalab³ a donc été lancée sur des serveurs en France en juin 2022.

Figure 3 : page d'accueil du datalab



Quatre services ont été proposés, issus des briques essentielles d'Onyxia (voir encadré ci-dessous). L'accès à ces services, hébergés sur une machine puissante, permet aux utilisateurs de charger leurs données en ligne, et de procéder à toutes les manipulations et tous les traitements qu'ils souhaitent, dans une optique de « test and learn », pour découvrir des outils à l'état de l'art de la science des données. En d'autres termes, il s'agit d'une véritable salle d'entraînement qui permet d'utiliser des bibliothèques de datascience (en R et Python) pour traiter de la donnée sans se préoccuper des enjeux liés à l'infrastructure (serveurs, puissance de calculs, etc.).

Encadré 1: présentation des services du datalab

Jupyter Notebook : outil permettant d'écrire et d'exécuter des scripts Python ou R au sein de documents.

R Studio : environnement de développement dédié au langage R. Permet d'importer des jeux de données et de développer des scripts R pour les analyser.

³ <https://datalab.datafid.world/>

2. Mise en place

2.1. Lancement du datalab en ligne

Dès lors que nous avons stabilisé les services inclus dans la plateforme, nous avons ouvert des accès aux partenaires du projet, et avons organisé des sessions de présentation et de prise en main. Par la suite, nous l'avons agrémenté de plusieurs fonctionnalités pour l'adapter aux réalités des pays et rendre son utilisation encore plus efficace.

D'abord, nous avons installé un composant open-source d'ETL (Extract Transform Load) : la structure des bases de données de certaines administrations partenaires est parfois ancienne et peu adaptée aux logiciels modernes de traitement des données. Par exemple, les bases de données de Sydonia en Côte d'Ivoire (système d'information des douanes, très répandu dans la région) sont basées sur une version dont le support date de 2011. Or, les langages et librairies de traitement de données ont fortement évolué depuis, rendant indispensable un connecteur : l'ETL. Il permet d'extraire des données à partir d'une véritable base de données en production, de les transformer et les adapter aux langages et librairies utilisées dans le datalab.

Ensuite, nous avons ajouté un mécanisme de « conteneurisation » avec Kubernetes (ou K8S dans le langage technique) : déjà intégré dans la logique d'Onyxia, il s'agit d'un système permettant d'automatiser l'infrastructure et lancer simultanément des machines virtuelles, ou conteneurs (dockers), qui agissent comme des ordinateurs indépendants et permettent ainsi à plusieurs personnes (voire un grand nombre) de travailler simultanément. Les opérations courantes d'accès, de maintenance et de démarrage des serveurs ne sont plus nécessaires, offrant ainsi un libre-service. Aucun serveur n'est surchargé et si une défaillance survient sur un serveur, l'anomalie est détectée et un redéploiement des conteneurs peut être assuré sans perte de données (en conformité avec les standards d'architecture big data). Des systèmes « d'étanchéisation » ont été prévus pour assurer la sécurité des données chargées sur le datalab.

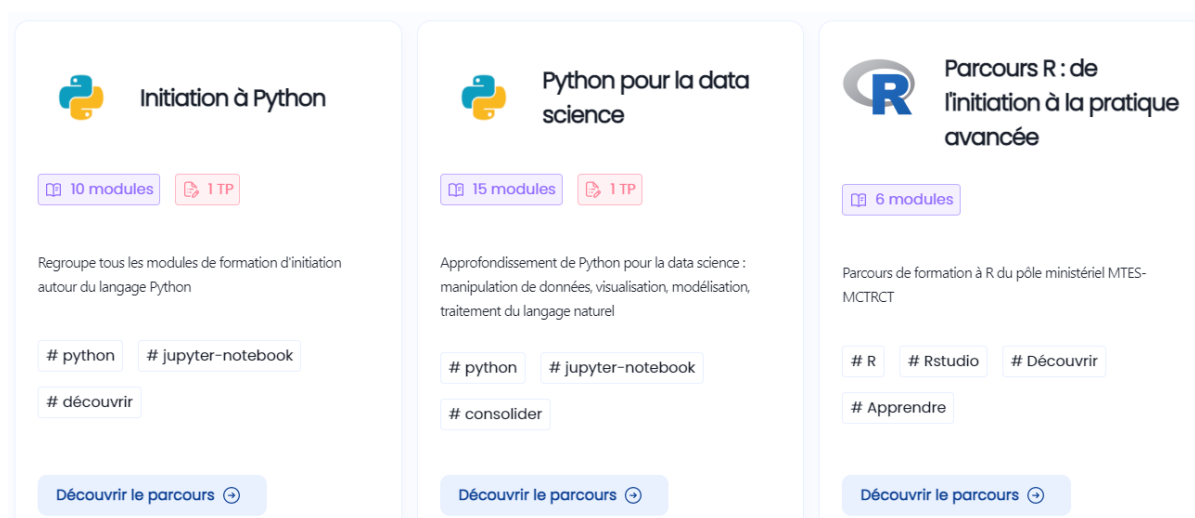
Enfin, et c'est l'une des dimensions clés de ce datalab, nous en avons fait un outil d'apprentissage et de formation : parmi les problèmes identifiés lors de la phase de



Réformes des finances publiques en Afrique francophone

démarrage, figurait le manque de compétences pour explorer le potentiel des données. Nous avons pris le parti de transformer le datalab en une librairie de formations accessibles aux partenaires titulaires d'un compte. Nous y avons chargé des vidéos de formations libres, issues de différents partenaires, que les utilisateurs peuvent suivre à leur rythme. Nous avons par ailleurs construit des parcours de formation qui permettent une montée en compétence progressive, depuis un niveau débutant jusqu'à un niveau confirmé.

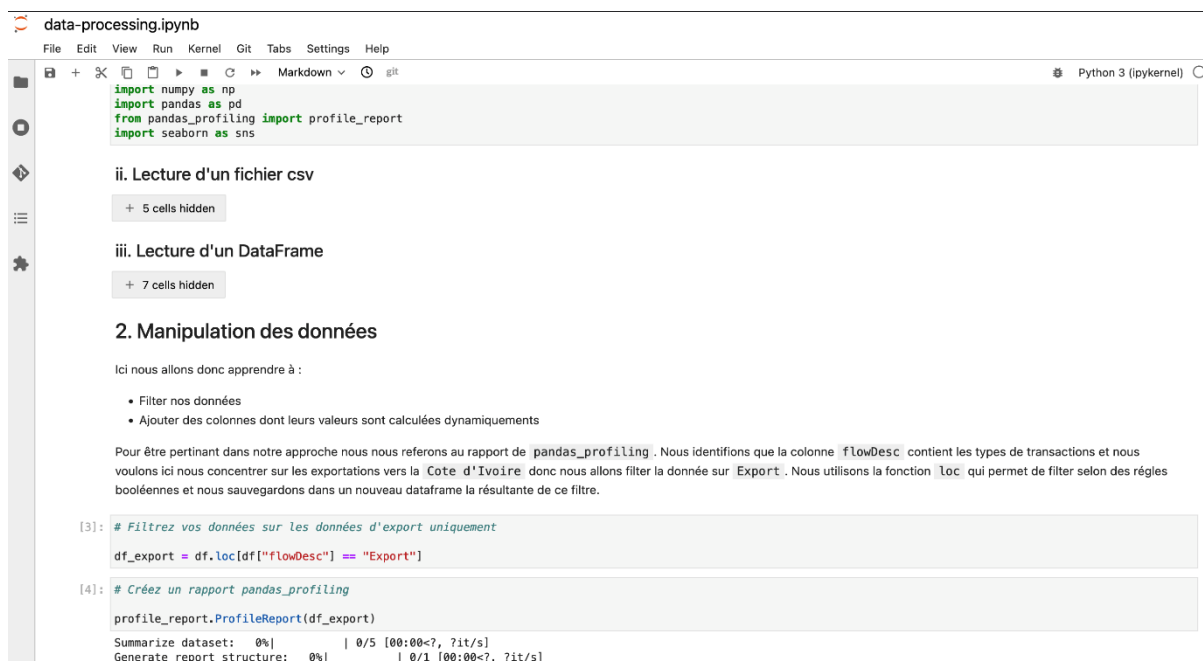
Figure 4 : vue des parcours de formation proposés sur le datalab



Enfin, nous avons construit un certain nombre de Travaux Pratiques (TP) directement en lien avec les métiers de la douane et des impôts. A titre d'exemple, nous disposons d'un TP qui explique comment appliquer la loi des nombres anormaux (loi de Benford) à des données douanières, pour identifier des anomalies dans certaines données statistiques.

Sur la période janvier – octobre 2023 nous avons enregistré environ 2000 usagers sur le datalab, qui comprend désormais près de 50 services, formations et TP.

Figure 5 : extrait du TP sur l'application de la loi de Benford aux données douanières



```
data-processing.ipynb
File Edit View Run Kernel Git Tabs Settings Help
Python 3 (ipykernel)
import numpy as np
import pandas as pd
from pandas_profiling import profile_report
import seaborn as sns

ii. Lecture d'un fichier csv
+ 5 cells hidden

iii. Lecture d'un DataFrame
+ 7 cells hidden

2. Manipulation des données

Ici nous allons donc apprendre à :
• Filter nos données
• Ajouter des colonnes dont leurs valeurs sont calculées dynamiquement

Pour être pertinent dans notre approche nous nous référons au rapport de pandas_profiling. Nous identifions que la colonne flowDesc contient les types de transactions et nous voulons ici nous concentrer sur les exportations vers la Côte d'Ivoire donc nous allons filter la donnée sur Export. Nous utilisons la fonction loc qui permet de filter selon des règles booléennes et nous sauvegardons dans un nouveau dataframe la résultante de ce filter.

[3]: # Filtrez vos données sur les données d'export uniquement
df_export = df.loc[df["flowDesc"] == "Export"]

[4]: # Créez un rapport pandas_profiling
profile_report.ProfileReport(df_export)
Summarize dataset: 0% | 0/5 [00:00<7, ?it/s]
Generate report structure: 0% | 0/1 [00:00<7, ?it/s]
```

2.2. Installation en Côte d'Ivoire

La Direction des Statistiques et des Etudes Economiques (DSEE) de la Direction Générale des Douanes de Côte d'Ivoire a souhaité bénéficier d'un déploiement du datalab au sein de son Système d'Information. Les arguments évoqués étaient comme suit :

- Les utilisateurs de la version en ligne doivent d'abord procéder à une anonymisation de leurs données avant de les charger sur le datalab. Un déploiement en ligne permettrait de se soustraire à cette étape et de travailler dans des conditions plus fluides, en attaquant directement Sydonia ;
- Le datalab en ligne est conçu pour réaliser des tâches exploratoires, pour s'entraîner. La DSEE a souhaité disposer d'un outil plus robuste, qui peut travailler sur des données de production, et proposer des analyses en temps réel ;
- La version en ligne du datalab comprend un certain nombre de services déterminés par le projet, selon les besoins exprimés. La DSEE a souhaité disposer de sa propre instance pour avoir la possibilité d'en ajouter d'autres, au gré des évolutions de son degré de maturité.

Réformes des finances publiques en Afrique francophone

En vue de réaliser ce chantier dans les meilleures conditions possibles, plusieurs éléments ont été envisagés, et plusieurs étapes respectées.

2.2.1. Portage politique

Un projet d'une telle envergure ne peut se faire sans l'assentiment du management de l'administration, et nécessite un portage politique au plus haut niveau. La DSEE a présenté son projet à la Direction Générale, et a inscrit le déploiement du datalab dans son plan d'action annuel, dont l'exécution fait l'objet d'un audit interne en fin d'exercice. La validation d'une telle activité lui donne un aval institutionnel important, dans la mesure où l'action implique la collaboration d'autres structures, au premier rang desquelles la Direction des Systèmes d'Information. Cette dernière, responsable de l'intégrité et de la sécurité de toute l'infrastructure de la douane, a dû être associée à la décision en exprimant les conditions dans lesquelles le déploiement pouvait avoir lieu. Une équipe multi-profil a été mobilisée pour participer aux actions de déploiement, formant un groupe de travail poursuivant un objectif commun.

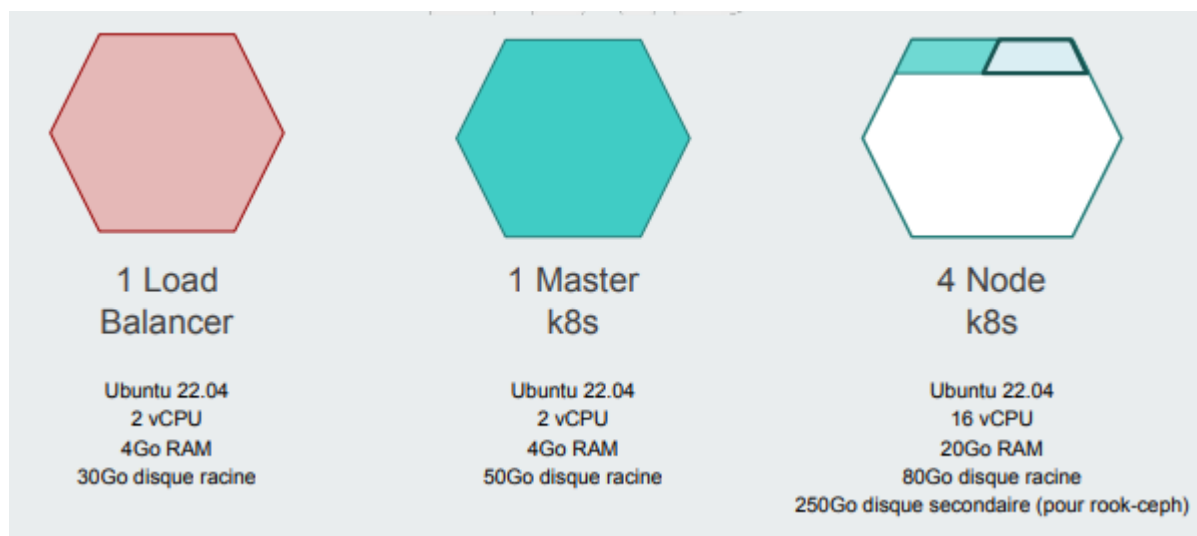
2.2.2. Actions techniques

Une fois les approbations politiques obtenues et les équipes constituées, des actions d'ordre technique ont pu commencer. Plusieurs experts ont été mobilisés par le projet (expertise en réseau et infrastructure, expertise en transition numérique) et ont réalisé un travail sur place et à distance. La première tâche a consisté en l'organisation d'une réunion de cadrage au profit de la DSI pour lui présenter dans le détail le fonctionnement de la version en ligne du datalab et expliquer les technologies mobilisées (Kubernetes, Docker) ainsi que les prérequis physiques eu égard à la volumétrie de données de la DGD (outils et puissances, composants, configurations).



Réformes des finances publiques en Afrique francophone

Figure 6 : prérequis techniques pour le fonctionnement du datalab



Dans un second temps, ont été conviés les collaborateurs des directions concernées, notamment la DSEE (à l'origine de la demande de déploiement), la Direction de l'Analyse Risque, du Renseignement et de la Valeur (DARRV) et la Direction des Enquêtes Douanières (DED), autres utilisatrices de données et « clientes » potentielles du datalab. La réunion a été l'occasion de baliser les attentes de chacun (nombre d'utilisateurs souhaités, utilisations prévues, implications en termes de circulation de la donnée) et d'identifier et lever tous les irritants inhérents à ce genre de projet (question liées à l'hébergement de la plateforme, à son administration, etc.). Enfin, la réunion a permis de préparer au mieux la visite sur place des experts en réseau et infrastructure, et notamment les conditions dans lesquelles ces dernières peuvent travailler. Les sujets ont notamment porté sur les ressources mobilisables (hyperviseur, réseau, CPU, RAM, disque), l'installation, la création et l'accès aux machines virtuelles, les noms de domaines utilisés, le schéma réseau et le plan IP des installations sur place, la politique de gestion des certificats SSL, l'ouverture sur l'extérieur pour accéder à d'autres jeux de données.

La mission sur place a été organisée en présence de plusieurs experts, qui ont œuvré à l'installation de la solution dans le système d'information et sa configuration en ligne avec ce qui a été identifié lors de la réunion de cadrage. La mission a abordé notamment les points de l'authentification, la question du stockage distribué, et un compte-rendu a été partagé aux équipes pour lister les étapes encore restantes.

Réformes des finances publiques en Afrique francophone

2.2.3. Conduite du changement

Le déploiement d'une solution ne doit pas s'arrêter à son installation physique ; il est fondamental d'accompagner les producteurs et utilisateurs de données pour qu'ils se saisissent pleinement des fonctionnalités de l'outil et qu'ils l'intègrent dans leurs tâches quotidiennes. Aussi, les équipes ont proposé plusieurs séances de travail avec des publics différents :

Avec la DSI, c'est-à-dire les « producteurs », une formation sur le maintien de la plateforme a été organisée dans la foulée du déploiement. Cette formation comprenait une présentation de l'architecture générale de la plateforme (Onyxia, KeyCloak, Minio), un retour sur les technologies de conteneurisation avec la technologie Docker, la découverte des principes d'orchestration de conteneurs avec Kubernetes, (performance, sécurité, résilience, exploitabilité), une présentation de l'architecture technique d'une solution kubernetes (noeuds master et workers, kubelet, api-server, proxy, dns, etc.). Une autre séance, sous forme de Travaux Pratiques (TP) doit avoir lieu, pour mettre les agents dans des situations concrètes et s'assurer qu'ils disposent des bons réflexes pour maintenir la plateforme. Par ailleurs, de la documentation technique a été produite et mise à disposition de la DSI pour qu'elle puisse s'y référer en cas de problème.

Avec la DSEE et la DARRV, c'est-à-dire les « utilisateurs », il a été proposé d'organiser un « sprint » de plusieurs jours pour passer en revue les fonctionnalités du datalab en utilisant les TP disponibles sur la plateforme (module de formation au langage Python notamment), puis de basculer sur la réalisation d'un exercice concret propre aux missions de la douane : la compilation d'analyses miroirs⁴. L'idée est de permettre aux équipes d'être autonomes sur le datalab afin qu'elles puissent réaliser leurs propres cas d'usage.

4

<https://www.wcoomd.org/fr/topics/research/~/~~/media/1FEA54015FAC46B0A0E72E73934329F0.a-shx>



Réformes des finances publiques en Afrique francophone

A noter par ailleurs que les équipes du projet demeurent disponibles jusqu'en avril 2024, date de fin des activités du projet.

2.2.4. Résultats obtenus

A l'heure de la rédaction de l'article (novembre 2023), le datalab est fonctionnel, mais l'exercice de conduite du changement est en cours ; il est un peu tôt pour tirer des leçons précises. Ce qui est certain néanmoins c'est que la présence de l'outil a permis à la DSEE de mieux envisager les façons dont elle peut remplir sa mission de production de statistiques et d'études économiques, et à la DSI de monter de façon significative en compétence sur des technologies à l'état de l'art.

3. Risques et difficultés associées

Les risques identifiés dans la mise en place de cette solution sont principalement de deux ordres : la question de l'appropriation et donc de la pérennisation de l'outil, et celle de la gouvernance.

3.1. Appropriation et pérennisation

Le datalab est basé sur des technologies entièrement open source. Les avantages de l'open source sont connus et tangibles : gratuité, maîtrise de l'outil et possibilité de le customiser à l'infini, mutualisation des connaissances avec une communauté d'utilisateurs, etc. Dans le cas présent, l'open source permet aussi de préserver la confidentialité des données : l'outil n'appartient à personne, et le chargement de données douanières, et donc sensibles, n'implique pas leur stockage dans l'infrastructure d'une entreprise privée. L'administration reste parfaitement souveraine.

Ces avantages ont toutefois un prix : un fort coût d'entrée. Le déploiement et surtout le maintien du datalab requièrent des compétences techniques avancées. On l'a vu en introduction, le diagnostic de maturité a souligné les faiblesses de ces compétences dans les administrations partenaires. La douane ivoirienne ne fait pas exception et des efforts importants en matière de formation doivent être consentis, sans quoi l'outil risque d'être sous-utilisé, voire abandonné, dès qu'il rencontrera un problème que les effectifs de la douane ne sauront résoudre.



Réformes des finances publiques en Afrique francophone

Par ailleurs, la montée en compétence permise par l'accompagnement du projet est vulnérable à la fuite des cerveaux. Les individus maîtrisant ces technologies de pointe peuvent être tentés de courtiser des emplois dans le secteur privé, souvent plus valorisés et mieux rémunérés.

3.2. La gouvernance des données

L'installation du datalab dans le SI de la douane ivoirienne est une réponse à une demande d'une Direction de disposer d'un outil lui permettant d'améliorer son travail. Toutefois, les conditions permettant d'atteindre cet objectif ne se limitent pas au renforcement des capacités des cadres pour qu'ils puissent maximiser le potentiel de l'outil. Il importe que l'administration dispose d'une gouvernance saine, c'est-à-dire un mode de fonctionnement qui permette aux moyens et ressources à disposition d'être utilisés de façon optimale pour atteindre l'objectif de l'organisation. Or, une gouvernance défectueuse peut générer des freins à deux niveaux : organisationnels et techniques.

Les problèmes d'organisation se caractérisent ici par une répartition sous-optimale des fonctions data. Dans le domaine de la douane, il existe plusieurs modèles d'organisation de ces fonctions, qui ne sont pas mutuellement exclusifs, mais qui ont des implications différentes :

- Un modèle type « datalab » ou « cellule data », qui centralise toutes les compétences data de la douane dans un service unique, et vers lequel se tournent les directions pour formuler leurs demandes et leurs cas d'usage. Ce service est alors responsable du développement de ces solutions, qui sont ensuite mises en production par les services informatiques ;
- Un modèle type « centre national de ciblage », qui mobilise plusieurs compétences data pour poursuivre un objectif unique, par exemple le ciblage des contrôles. Il s'agit d'un modèle « transitionnel », une étape par laquelle les administrations souhaitant devenir « data-oriented » passent, pour concentrer des ressources dans un secteur clé et chercher à monter en compétence sur ce périmètre avant d'investiguer de nouveaux champs ;



Réformes des finances publiques en Afrique francophone

- Un modèle type « training center » qui forme les douaniers à la science des données avant leur dissémination dans toutes les directions susceptibles de mobiliser de la donnée.

Il ne s'agit pas de choisir entre l'un ou l'autre des modèles, mais simplement de réfléchir à ce qui est attendu des données dans l'administration et de décider en conséquence de l'organisation la plus appropriée. Dans le contexte ivoirien, cette réflexion est balbutiante et de nombreux chevauchements existent entre plusieurs directions, notamment DSEE, DSI et DARRV : la DSEE dispose d'un modèle économétrique d'analyse de risque et produit des statistiques douanières, la DSI administre Sydonia et exploite ses données dans un modèle de Business Intelligence pour fournir des indicateurs de pilotage, et la DARRV possède son modèle de sélectivité des contrôles a priori. Ces chevauchements, qui témoignent d'un enthousiasme pour la question de la donnée qui ne doit être bridé, gagneraient à être rationalisés. Le déploiement d'un outil type datalab, s'il permet d'appuyer l'une ou l'autre des directions concernées, ne saurait ni résoudre la question de la répartition des fonctions données, ni régler le problème des manques à gagner que génèrent ces chevauchements.

Les problèmes de gouvernance peuvent se faire ressentir sur des choix techniques et concerner la gestion de la donnée elle-même. Ce sont généralement le résultat de dysfonctionnements institutionnels et organisationnels, qui reflètent l'absence de vision globale sur ce que l'on souhaite faire de la donnée comme actif. Ils se manifestent par des problèmes plus « micro », ou « techniques », comme par exemple le maintien d'un système d'information peu performant, qui repose sur l'utilisation d'outils anciens et peu adaptés. Dans le contexte ivoirien, c'est le cas notamment de la base qui héberge les données Sydonia, qui utilise une ancienne technologie (version Oracle de 2011), peu propice à la mobilisation de nouvelles techniques comme le machine learning. Ils peuvent aussi se caractériser par des problèmes dans le traitement et le partage des données. A titre d'exemple, jusqu'à très récemment, la DSEE n'avait accès qu'à certaines données de Sydonia, et souffrait d'un manque de visibilité sur la composition des champs additionnels du système, ce qui empêchait toute réflexion sur des nouveaux usages permis par les données. Enfin, ils impactent la qualité de la donnée, qui dépend de la célérité des



Réformes des finances publiques en Afrique francophone

agents qui remplissent les déclarations. Parmi les grandes missions douanières⁵, l'administration ivoirienne se focalise principalement sur la mission fiscale, à savoir la collecte des droits et taxes. Par conséquent, dès lors qu'une opération n'impacte pas la mobilisation des ressources intérieures (export d'une marchandise non taxée par exemple), l'incitation à renseigner les champs de façon précise et exhaustive est moindre, faisant peser la menace d'une donnée au mieux incomplète, et souvent fautive. Comme le dit l'adage, « garbage in, garbage out » : le datalab n'est qu'un outil dont les performances dépendent de la qualité de la matière première (la donnée) qu'il exploite.

4. Perspectives et recommandations

S'il est trop tôt pour observer d'éventuels changements induits par l'arrivée du datalab dans les performances organisationnelles et fiscales de l'administration douanière, il est toutefois possible de discuter des perspectives pour la suite, qu'il s'agisse du déploiement en Côte d'Ivoire ou de nouveaux développements de la version en ligne du datalab, et de dresser quelques recommandations qui permettent notamment de mitiger les risques ci-dessus identifiés.

4.1. Renforcer la conduite du changement avec davantage de formations, et davantage de cas d'usage

L'un des risques identifiés portait sur la non pérennisation de l'outil déployé dans l'administration douanière ivoirienne. Ce risque se matérialise notamment par un non usage de la solution, qui pourrait être vue comme trop complexe. La solution réside dans l'atteinte d'un seuil de résilience d'ici la fin du projet, actée au 30 avril 2024, pour d'une part baisser le coût de sa maintenance, et d'autre part rendre l'abandon de la solution particulièrement dommageable du fait de sa pleine intégration dans le système et des avantages qu'il procure aux utilisateurs.

Ce seuil doit pouvoir être atteint de plusieurs façons. Tout d'abord, de nombreuses séances de présentation du fonctionnement de la solution au profit de la Direction des

⁵ Fiscalité, protection du territoire et facilitation des affaires.



Réformes des finances publiques en Afrique francophone

Systèmes d'Information ont été organisées. A la date de rédaction de l'article, d'autres séances sont prévues, et le projet s'active par ailleurs à la production d'une documentation exhaustive à laquelle les équipes pourront avoir recours en cas de problème. La prochaine étape consistera en l'intégration de l'équipe de la DSI dans la communauté utilisatrice du datalab, animée par l'INSEE et qui rassemble de plus en plus de structures convaincues par la puissance de l'outil. Le recours communautaire est un puissant moyen d'apporter une solution à un problème rencontré par l'un des usagers. Des forums d'échange existent déjà, et des réunions sous forme de retours d'expérience et de présentation des usages de chacun se tiennent régulièrement. Faciliter l'accès à l'information et au « débogage » doit permettre de contribuer à atteindre ce seuil de résilience, en baissant le coût de sa maintenance.

Il convient ensuite d'atteindre ce seuil de résilience en faisant la démonstration de la puissance de l'outil pour le rendre indispensable. Cela passe par une poursuite de l'accompagnement aux utilisateurs afin qu'ils continuent à développer des cas d'usage pertinents et fondamentalement utiles à l'administration, appliqués à plusieurs fonctions de l'administration, principalement en lien avec la fonction fiscale (reporting des recettes journalières, analyse risque, déploiement de techniques pour accélérer les fonctions d'évaluation de la valeur en douane) mais qui concernent aussi des éléments plus transversaux qui concernent des aspects de gestion des ressources humaines (amélioration de la gestion calendaire, introduction de contrats de performance) ou tout simplement l'automatisation de certaines tâches chronophages (création de scripts pour des analyses statistiques automatiques).

4.2. Proposer des nouveaux services et étendre à de nouvelles directions

La stratégie d'implantation de l'outil et de sa pérennisation passe aussi par un développement constant de la solution en ligne, et par une extension de son utilisation à d'autres administrations.

La version en ligne du datalab, maintenue par le projet, est personnalisable à l'infini. Le récent fulgurant déploiement des outils d'intelligence artificielle générative constitue l'une des pistes sur lesquels nous travaillons : nous testons actuellement dans la version en ligne un module dans ce sens. Si les essais s'avéraient concluants,



Réformes des finances publiques en Afrique francophone

il serait alors possible de déployer ce module dans la version installée en Côte d'Ivoire, pour contribuer à offrir de nouvelles opportunités aux jeunes cadres et techniciens formés à ces technologies.

Par ailleurs, l'approche communautaire précédemment mentionnée pourrait être nationalisée ou régionalisée, c'est-à-dire pensée aussi dans le contexte ivoirien ou sous-régional, en déployant la solution dans d'autres administrations. Des bribes de discussion dans ce sens ont eu lieu avec d'autres administrations et d'autres pays de la région, sans pour autant atteindre le même degré d'avancement qu'avec la douane ivoirienne. L'implantation de l'outil dans d'autres structures viendrait toutefois renforcer le leadership de la douane ivoirienne en la matière, et contribuer à renforcer son utilisation dans d'éventuels autres services que ceux initialement visés.

4.3. Fidéliser les collaborateurs formés

La démultiplication des cas d'usage possibles et l'introduction de nouvelles technologies constituent une des façons de fidéliser les collaborateurs formés. L'accès aux technologies proposées pour le datalab, le volume de données disponibles, et la mission de service public utile à la société que remplit l'administration douanière constituent un formidable terrain de jeu pour des jeunes cadres passionnés par l'apprentissage et soucieux du bien collectif. Les résultats obtenus sont la meilleure façon de valoriser le travail des personnels en charge du traitement de la donnée, valorisation à même de contribuer à la fidélisation des cadres concernés. C'est cela dit une condition nécessaire mais non suffisante à leur rétention, et la fidélisation des ressources humaines doit passer par la production d'une stratégie globale, et donc s'attaquer à la question de la gouvernance.

4.4. Adopter une stratégie pour devenir une organisation orientée vers les données

La gouvernance, ou plutôt les dysfonctionnements de cette dernière, a précédemment été mentionnée comme étant une cause de la non-atteinte des objectifs du datalab. Le projet DATAFID s'est caractérisé par son côté exploratoire : il a permis de mieux saisir les degrés de maturité des administrations concernées, de détecter les talents à mobiliser pour la production de cas d'usage, et d'identifier les zones d'amélioration



Réformes des finances publiques en Afrique francophone

dans le cycle de la donnée. Cette approche a permis une entrée par le bas, du côté technique.

Le déploiement du datalab est une amorce de remontée vers des aspects plus stratégiques car elle demande que l'administration une réflexion sur ce qu'elle souhaite faire de ses données. Cette amorce doit être concrétisée : il convient désormais d'adopter une approche plus large, plus systémique, en d'autres termes, d'embrasser pleinement la transition numérique pour devenir une administration orientée vers la donnée. De cette réflexion découlera une vision stratégique pour l'ensemble de l'administration douanière, dont la mise en œuvre sera certainement porteuse de changements dans l'organisation des fonctions data, et donc certainement des mutations dans les pratiques de gestion des ressources humaines (profils à recruter, gestion des carrières, formation continue, etc.). Il s'agit là d'un chantier de taille aux implications politiques, que le seul projet DATAFID ne pourra pas porter. Les équipes ont toutefois débuté des actions allant dans ce sens pour baliser le sentier de la réflexion et commencer à poser quelques jalons :

- En décembre 2022, le projet a mandaté une mission de l'Organisation Mondiale des Douanes, qui a rencontré les acteurs principaux de la douane et a passé l'organisation des fonctions données. Quelques recommandations ont été proposées pour améliorer notamment la circulation des données, recommandations qui ont permis à la DSEE d'avoir accès à davantage de données produites par la DSI.
- Au premier semestre 2022, le projet a produit un livre blanc intitulé « Comment les administrations fiscales et douanières peuvent améliorer leurs performances grâce à la donnée ? », qui a été remis aux points focaux du projet, y compris au sein de la douane ivoirienne. Ce livre blanc contient une sorte de recette de la transition numérique, qui doit bien sûr être contextualisée avant d'être déployée, et qui a pour but de lister les grandes étapes que les administrations peuvent suivre pour opérer cette mutation avec succès.
- En juillet 2023, le projet a mandaté une autre mission de l'OMD pour une rencontre avec le management de la DGD. La rencontre a permis de discuter des recommandations formulées par le FMI lors d'une mission du fonds effectuée fin 2022. Il est apparu que les recommandations étaient trop larges et peu



Réformes des finances publiques en Afrique francophone

opérationnelles, et que des approfondissements étaient nécessaires. La DGD a exprimé son besoin d'une assistance dont les contours doivent être discutés.

- Parmi les propositions exprimées par l'équipe projet, figure l'organisation d'un séminaire au T1 2024 rassemblant les directeurs des directions concernées par la production et l'utilisation de données, accompagnées d'experts en transition digitale et de douaniers, qui doit permettre d'aboutir à un consensus autour de la vision de la donnée, première pierre nécessaire à la création d'une stratégie data.
- Le projet prendra fin en avril 2024 mais les réussites de la phase en cours alimentent les contours d'une prochaine phase. La vision proposée par l'équipe est de basculer d'une approche « projet » à un « produit », ou initiative non limitée dans le temps, comportant une offre de service complète et intervenant sur demande, un peu comme le dispositif « Inspecteurs des Impôts Sans Frontières » de l'OCDE⁶. Un des grands axes concentrerait des ressources sur un appui à une transition vers des administrations orientées vers la donnée.

⁶ <http://www.tiwb.org/fr/>

