

Projet d'Appui au Renforcement des Capacités Statistiques (ARCS)

MISSION INSD 5
«CONCEPTION D'UN CAHIER DES CHARGES GÉNÉRAL DES SYSTÈMES DE GESTION
DES BASES DE DONNÉES DE L'INSD»

13 avril au 5 mai 2007

Rapport final

Sommaire

| | |
|--|----|
| Liste des abréviations et acronymes | 4 |
| Résumé | 5 |
| 1 Introduction | 10 |
| 1.1 Objectifs de la mission | 10 |
| 1.2 Organisation des travaux de la mission | 10 |
| 2 Le système d'archivage actuel des données de l'INSD | 11 |
| 2.1 Introduction | 11 |
| 2.2 Description du système actuel d'archivage des données à l'INSD | 12 |
| 2.2.1 Les moyens informatiques de l'INSD | 12 |
| 2.2.2 La sécurité des données | 14 |
| 2.2.3 L'archivage des données des opérations statistiques | 15 |
| 2.2.4 La base de données socio économiques et les principales bases de données statistiques | 17 |
| 2.2.5 Les projets passés, en cours et prévus dans le domaine | 18 |
| 2.3 Diagnostic du système actuel | 19 |
| 3 Le futur système d'archivage électronique des données de l'INSD | 20 |
| 3.1 Les objectifs | 20 |
| 3.2 Présentation des fonctionnalités du SAE de l'INSD | 20 |
| 3.2.1 Détail fonctionnel du SAE de l'INSD | 21 |
| 3.3 Proposition d'architecture générale du SAE | 25 |
| 3.4 La sécurité des données – Plan de sécurité informatique | 26 |
| 3.4.1 Priorités de sécurité à court terme | 26 |
| 3.4.2 Plan de sécurité des données | 26 |
| 3.5 Archivage des données des opérations statistiques | 27 |
| 3.5.1 L'information à archiver | 27 |
| 3.5.2 Les fonctions à assurer | 28 |
| 3.5.3 Les moyens | 30 |
| 3.5.4 Les étapes de mise en œuvre | 31 |
| 3.6 La base de données socio-économiques | 32 |
| 3.6.1 Les informations à stocker dans la base de données socio-économiques | 33 |
| 3.6.2 Les fonctions assurées par le système de gestion de la base de données socio- économiques | 33 |
| 3.6.3 Les moyens | 35 |
| 3.6.4 La mise en œuvre | 36 |
| 4 Plan d'action prioritaire | 37 |

| | | |
|-----|--|----|
| 4.1 | Objectif 1 : Sécuriser les données traitées..... | 37 |
| 4.2 | Objectif 2 : Archiver les données des dernières enquêtes réalisées et des opérations statistiques annuelles stratégiques | 37 |
| 4.3 | Objectif 3 : Construire une première version de la base de données socio-économiques..... | 38 |
| 5 | Conclusion..... | 39 |
| | Annexes..... | 40 |
| | Annexe 1 - Description des principales bases de données..... | 40 |
| | Annexe 2 – Les outils et projets d’archivage des données..... | 42 |
| | Le projet Databank..... | 42 |
| | Le Toolkit pour l’archivage des données | 42 |
| | Le projet ADP (Programme statistique accéléré pour l’Afrique) | 42 |
| | Le projet IMIS (Système intégré de gestion de l’information) | 44 |
| | Annexe 3 – Principales solutions existant en matière d’entrepôts de données statistiques . | 45 |
| | Beyond 20/20 | 45 |
| | DevInfo | 46 |
| | PC-Axis | 47 |
| | REDATAM+SP | 49 |
| | Annexe 4 - La mise en œuvre de l’objectif 1 du programme prioritaire..... | 50 |
| | 1. Estimation des coûts pour la mise en oeuvre | 50 |
| | 2. Schéma de mise en œuvre (Feuille de route) | 51 |
| | Annexe 5 - Liste des personnes rencontrées | 52 |

Résumé

Conformément aux termes de référence, les travaux de la mission ont consisté, dans un premier temps à faire un diagnostic du système actuel de stockage et d'archivage des données de l'INSD et, dans un second temps à proposer un cahier des charges des systèmes de gestion des bases de données de l'INSD comprenant un plan d'action pour la mise en œuvre du système d'archivage des données.

Le premier volet de la mission s'est appuyé sur des entretiens avec les responsables des services chargés de la collecte, du traitement et de la diffusion statistique au sein de l'INSD. Au cours de cette phase il est apparu très vite que la sécurité et la pérennité des données de base et des résultats présentaient de telles lacunes qu'elles étaient de nature à compromettre tout système d'archivage des données. La mission a donc inclus dans son champ de travail cet aspect. Le second volet de l'étude s'est centré sur l'élaboration du cahier des charges des systèmes de gestion de base de données et d'un plan d'action à court terme.

1. Évaluation du système d'archivage de l'INSD

D'une façon générale, le système d'archivage actuel des données statistiques au sein de l'INSD est extrêmement précaire. Des ensembles importants de données sont menacés de disparition ou d'altération si des mesures pour assurer la sauvegarde courante des données et leur sécurisation ne sont pas prises très rapidement.

Le personnel est encore peu sensible aux enjeux liés à la sécurité des données et à leur archivage. L'infrastructure informatique existante ne fournit pas aujourd'hui tous les services indispensables pour sécuriser les données présentes sur les postes de travail, en particulier pour garantir leur intégrité et leur protection contre des virus. Ceci est en partie la conséquence de l'absence d'un réseau local, absence qui ne peut être justifiée aujourd'hui par une insuffisance des moyens matériels existants. L'insuffisance des moyens humains du Service informatique a conduit à des arbitrages qui privilégient les appuis aux opérations ponctuelles au détriment d'une assistance permanente aux utilisateurs comme cela serait nécessaire dans le cadre du fonctionnement d'un réseau local.

Un des points forts de l'INSD est sa pratique d'archivage des données d'enquêtes auprès des ménages depuis plus de dix ans maintenant. Toutefois cet archivage est encore peu structuré et il doit être étendu à tous les types d'enquête et aux données produites à partir des dispositifs administratifs.

L'absence d'une base de données socio-économiques constitue un handicap pour produire des publications générales comme l'Annuaire statistique, pour alimenter le site internet et pour répondre aux demandes les plus fréquentes des utilisateurs.

2. Les principaux éléments du cahier des charges des systèmes de gestion des bases de données de l'INSD

Le cahier des charges porte sur les trois composantes du système d'archivage :

- La sécurité des données.
- L'archivage des données de base.
- La base de données socio-économiques.

Le schéma général proposé est cohérent avec celui retenu par le schéma directeur informatique de l'INSD. Pour les 2 dernières composantes, l'architecture retenue s'appuie sur celle proposée par le modèle de référence OAIS (Open Archival Information System) qui

identifie 5 fonctions : versement, stockage, gestion des données descriptives, consultation et communication, administration.

1. Composante 1 : Sécuriser les données

Les données à sécuriser sont celles présentes sur les postes de travail, et à l'avenir sur les serveurs. Les fonctions à réaliser par le système sont :

- La sauvegarde régulière et automatisée des données par un dispositif permettant de les restaurer rapidement avec un minimum de pertes sur les données créées les plus récentes.
- Le contrôle des accès aux postes de travail et aux données.
- La protection contre toute forme de destruction ou d'altération par les virus.

Les moyens matériels et outils logiciels nécessaires sont déjà pour l'essentiel présents au sein de l'INSD. Il s'agit de mettre en place un réseau local en utilisant le serveur existant et en y déployant les services assurant les différentes fonctions ci-dessus. Quelques investissements complémentaires assez peu coûteux sont à prévoir pour mettre à jour le logiciel du serveur et consolider les équipements informatiques permettant de connecter les réseaux Wifi du site central et du bâtiment annexe. Les priorités données au Service informatique doivent être revues pour assurer une permanence du fonctionnement du réseau et l'assistance aux utilisateurs. La participation active du personnel doit être stimulée par une formation/information sur la sécurité des données.

La mise en œuvre de ces nouvelles fonctions peut être très rapide (cf. plan d'action).

2. Composante 2 : Archivage des données de base

L'archivage des données de base concerne les données, les métadonnées et les résultats produits par l'ensemble des opérations statistiques : enquêtes, exploitation des données issues de sources administratives et élaboration de données de synthèse (comptes nationaux et tableaux de bord).

Les fonctions à assurer par le système sont les suivantes :

- Collecte des données à archiver.
- Stockage des données archivées.
- Administration des archives.
- Consultation des archives.

L'INSD dispose déjà des moyens logiciels pour assurer ces fonctions avec le Toolkit de la Banque mondiale et la base de données des archives installée dans le cadre du projet Databank. Le Service Informatique doit être en mesure d'assurer la centralisation et la gestion de ces archives, ce qui ne représente pas en soi une charge annuelle de travail très importante, mais une constance dans la fourniture du service. De même la prise en compte de cette préoccupation par les services producteurs n'est pas une charge très importante, pour autant que cette tâche soit prise en compte tout au long du déroulement de l'opération statistique. Des formations des statisticiens à l'utilisation du Toolkit sont également à prévoir.

La mise en œuvre de l'archivage des données de base à l'INSD passe par plusieurs étapes :

- La définition par la Direction générale de l'INSD des procédures et rôles des différents acteurs impliqués dans cette opération et de la politique de diffusion des archives.
- La mise en place des outils et des moyens humains.

- La prise en compte de l'archivage dans les opérations en cours et l'archivage des données récentes.
- La reprise des données déjà archivées pour les mettre dans le format « Toolkit » et générer de nouvelles versions des supports de stockage.
- La confection des archives des opérations plus anciennes : elle nécessite un inventaire des données disponibles et se fera selon un calendrier qui tiendra compte de ces disponibilités et du degré d'urgence face aux situations de stockage des données existantes.

3. Composante 3 : Base de données socio-économiques

C'est un dispositif complémentaire à l'archivage des données de base et aux bases de données localisées comme le système IMIS. Il concerne les données statistiques les plus couramment utilisées et demandées, ainsi que les métadonnées correspondantes.

Les fonctions à assurer par le nouveau système sont les suivantes :

- Alimentation de la base de données
- Chargement des données
- Chargement des métadonnées
- Consultation de la base de données
- Administration de la base de données

De telles bases de données utilisent généralement des solutions logicielles spécifiques comme Beyond 20/20, PC-Axis, DevInfo ou IMIS. L'INSD devra faire un choix entre ces solutions ayant fait leurs preuves en tenant compte de leurs spécificités techniques au regard de ses besoins, mais aussi des possibilités d'appui offertes par leur promoteur. La mise en place d'une telle base nécessite des moyens humains importants dans la phase de déploiement, mais également des moyens de gestion permanents du système en phase courante.

La mise en place de cette base de données se situe dans une perspective à moyen terme (3 à 5 ans). Elle passe par les étapes suivantes :

- Une solution provisoire à court terme (cf. le plan d'action).
- Le choix de l'outil logiciel utilisé comme support.
- L'élaboration du projet de mise en place de l'outil.
- La mise en place progressive de la base de données couvrant d'abord les productions de données statistiques de l'INSD, puis celles des autres services du SSN.

3. Le plan d'action

Ce plan a pour objectif de mettre en œuvre dans les 12 mois à venir des actions prioritaires répondant aux besoins les plus urgents identifiés au cours de la mission.

Objectif 1 : Sécuriser les données

Action : Mettre en place un réseau local fournissant les services de sauvegarde et restauration des postes de travail, gérant l'accès aux postes et aux données, et assurant la diffusion d'un antivirus maintenu à jour.

Moyens : Un appui au Service informatique par une société externe permettant un déploiement rapide du réseau local et un transfert de compétences au service informatique.

Quelques compléments en matériels

Coût estimé : 13 millions FCFA

Ressources humaines : mise en place au Service informatique d'une équipe permanente de 2 personnes (1 informaticien, 1 technicien ayant des compétences en gestion de réseau et système d'exploitation Windows SBS 2003 server).

Objectif 2 : Archiver les données des dernières enquêtes et des opérations statistiques stratégiques

Actions :

1. Organiser l'archivage (procédure et rôles, politique de diffusion)
2. Mettre en place le répertoire des archives (adaptation de la base de données fournie par le projet Databank)
3. Formation à l'utilisation du Toolkit
4. Archiver : QUIBB (2007, 2005), MICS (2006), EDS (2003), Comptes nationaux, Commerce extérieur, IHPC

Moyens :

Renforcement du Service informatique pour lui permettre d'assurer sa fonction d'administration de l'archivage

2 mois*personne d'informaticien pour la mise en place de l'archivage
0,5 mois*personne d'informaticien et 0,5 mois*personne de technicien par an pour la gestion de l'archivage.

Appui du projet ADP dans le cadre de sa phase 1 (archivage des données)
Appui du projet ARCS pour la mise en place du répertoire des archives (2 mois*informaticien)

Objectif 3 : Construire une première base de données socio-économiques

Action : Développer une base de données de tableaux statistiques et la mettre sur le réseau.

Moyens :

Valoriser les données existantes qui sont souvent disponibles sous la forme de tableaux Excel

Un statisticien pendant une année pour le chargement et le formatage des tableaux.

Appui du projet ARCS pour le développement du système d'interrogation et de restitution de la base de données de tableau et pour l'automatisation de la production des tableaux de l'Annuaire statistique (6 mois*informaticien).

4. Conclusion

La grande précarité de la situation des données traitées et produites par l'INSD a fait passer un peu en second plan les objectifs centraux de la mission sur l'archivage électronique des données. Pour garantir une sécurité minimale aux données traitées et produites, et favoriser la diffusion des résultats, la mission propose un plan d'action à court terme. Le cahier des charges des systèmes de gestion des bases de données qui est proposé se place plus dans une perspective de moyen terme.

Le Service Informatique de l'INSD est au cœur de toutes ces problématiques. Si les missions qui lui sont attribuées par les textes organisant l'INSD n'appellent pas de remarque, ses effectifs et son fonctionnement actuel doivent être réexaminés dans les meilleurs délais. En effet, si le réseau local devient l'axe central pour assurer la sécurité des données et leur archivage, le service qu'il va devoir assurer aux utilisateurs exige une permanence qui n'est pas compatible avec les priorités qui lui sont données actuellement. La mise en place d'une base de données socio-économiques va renforcer cette exigence de permanence dans le

moyen terme. Le renforcement du Service Informatique doit se faire à la fois par des informaticiens mais aussi par des cadres statisticiens intéressés et motivés par le domaine.

Le recours à la sous-traitance est proposé dans le cadre du plan d'action comme un appui ponctuel au Service informatique, lui permettant de consolider ses compétences. Mais dans le cadre d'une réflexion globale sur les effectifs et les priorités données aux travaux qui lui sont confiés, la sous-traitance de la gestion du réseau et de l'assistance aux utilisateurs devrait faire partie des solutions explorées.