



n°9519/DT

RAPPORT DE SYNTHESE METHODOLOGIQUE SUR LA SAISIE DES

TROIS PHASES DE L'ENQUETES 1-2-3

ET DU RECENSEMENT INDUSTRIEL 1995

Janvier 1996 ⁽¹⁾

Le projet MADIO se propose d'apporter une contribution à la réflexion économique et sociale sur le processus de transition à Madagascar. Pour mener à bien cet objectif des enquêtes statistiques ont été effectuées. Il s'agit des enquêtes 1-2-3 (enquêtes sur les conditions d'emploi, des conditions de vie et du secteur informel d'Antananarivo), du Recensement Industriel, de la mise en place d'observatoire en zone rurale et les indices de prix à la consommation.. La présente note fournit les méthodologies appliquées pour **la saisie informatique des informations collectées et la mise en place de la base de donnée finale** (fichiers apurés) lors des enquêtes 1-2-3 et du recensement industriel. Etant donné le fait que certaines de ces enquêtes aient été effectuées pour la première fois à Madagascar et vu les contraintes matérielles de l'INSTAT, des problèmes sont apparus lors du déroulement des opérations informatiques. Ils ont pu être néanmoins résolus et serviront à coup sûr de leçons pour les enquêtes à venir. Ainsi, nous essayerons de donner le maximum d'information possible en annexe sur les " dessous de l'informatique " afin qu'il y ait un rendement d'échelle dans l'informatisation des enquêtes menées dans le cadre du Projet MADIO en particulier et de l'INSTAT en général.

Le plan du rapport est le suivant : dans un premier temps, nous essayons de mettre en exergue l'environnement dans lequel le travail d'informatisation s'est effectué. Les différentes contraintes qui ont conduit au choix de logiciel utilisé et du lieu de saisie y sont tout particulièrement mentionnées. Dans un second temps, le schéma conceptuel de la modélisation est présenté avec les quantifications temporelles de chaque étape. A cet effet, un tableau récapitulatif de la durée de chaque étape est fourni dans cette partie. Nous proposons ensuite quelques améliorations pour l'informatisation des futures enquêtes compte tenu des expériences déjà acquises. Pour clore ce rapport, un paragraphe est réservé à l'explication des annexes.

¹⁾ Ce document a été rédigé par Jean-Gabriel RANDRIANARISON, projet MADIO.

I.- L'ENVIRONNEMENT INFORMATIQUE DE L'ENQUETE

Des contacts préliminaires ont été effectués auprès des directions de l'INSTAT qui ont déjà eu des expériences en matière de traitement informatique des informations. Il s'agit notamment de la Direction du Recensement de la Population et de la Direction de l'Exploitation des Systèmes et Assistance à l'Informatisation (DESAI). Au cours des entretiens qui ont eu lieu, les responsables de chacune de ces directions nous a expliqué leur méthode de travail et les différents matériels qui sont à leurs dispositions. Les contraintes en terme de charge de travail et de personnel disponible ont été aussi évoquées. Ainsi, la Direction du Recensement de la Population nous a fait part de la non disponibilité de leurs opérateurs de saisie sauf en dehors des heures ouvrables.

Le choix du logiciel

Trois choix possibles se sont présentés pour la saisie des données quant aux logiciels disponibles : le logiciel IMPS, le logiciel PERFORM PRO, le logiciel DBASE . Chacun de ces trois logiciels ont été décortiqués afin d'estimer leurs possibilités et leurs faiblesses respectives. Une fiche technique de chacun d'eux sont fournis en annexe. Le masque de saisie de la DESAI (développé sous DBASE et CLIPPER) a été finalement retenu. **Celui-ci présente l'avantage de la double saisie et de contrôle de champ. Par ailleurs, les opérateurs de saisie ont déjà une facilité de manipulation de ce logiciel.** L'intégration des test de cohérence dans le masque de saisie n'a pas pu être réalisée à cause de sa trop grande rigidité.

Le lieu de la saisie

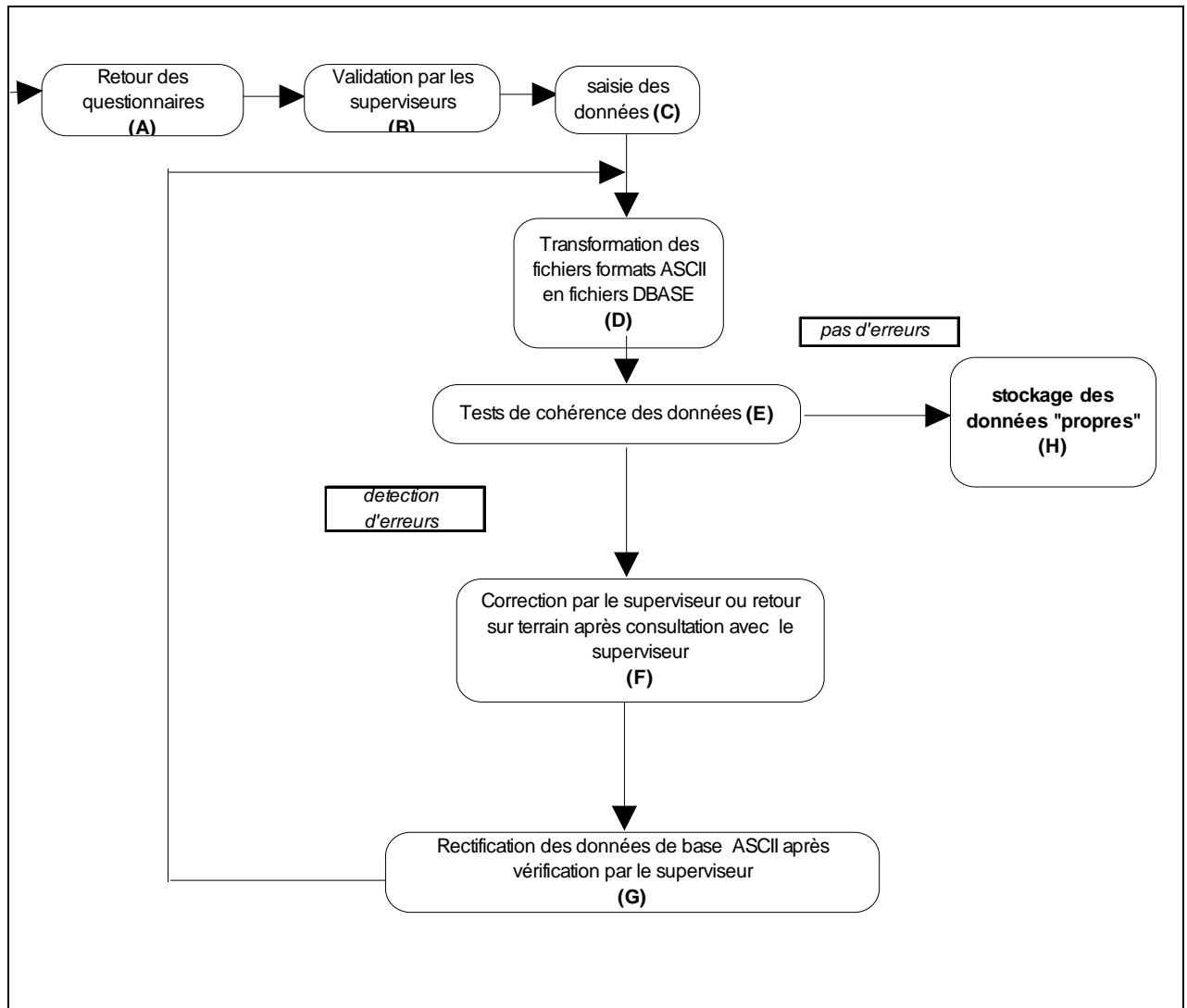
La saisie a été effectuée à l'INSTAT à la DESAI. Ce choix repose sur des raisons pratiques dans la mesure où le projet MADIO est implantée à la Direction de la statistique même et que le masque de saisie est celui de la DESAI. Néanmoins, la vétusté des matériels et les charges de travail des opérateurs de saisie (solde des fonctionnaires, pension des retraités, bulletin électorale) ont quelques peu ralenti les travaux. L' alternative aurait été de faire la saisie des données à la Direction Générale du Recensement qui dispose de matériels plus neufs et d'un logiciel spécialisé dans le traitement des enquêtes mais étant donné l'éloignement géographique et la charge de travail du personnel de recensement, cette option a été abandonnée.

II.- LES DIFFERENTS NIVEAUX DE PRESENTATION DU SCHEMA INFORMATIQUE

Les questionnaires retournés à l'INSTAT et validés par les superviseurs sont groupés par " vague". Chaque vague doit passer à travers la machine de modélisation du Schéma I jusqu'à apurement total des erreurs. Le nombre de vague utilisé varie suivant la taille du questionnaire.
(Voir Tableau III).

SCHEMA I

Schéma conceptuel du modèle informatique



Détail de chaque étape du modèle

Etape A

Les questionnaires sont remis aux superviseurs .

Etape B

Validation des questionnaires par les superviseurs.

Etape C

Les questionnaires validés sont groupés par vague et envoyés à la salle informatique pour y être saisis.

Etape D

Les formats des fichiers étant en ASCII, un programme de traitement les transforme en fichier DBASE. Cette étape est connu sous le nom d'éclatement.

Etape E

Des tests de cohérence seront effectués sur les fichiers DBASE. Les détails ceux-ci sont présentés an annexe.

Etape F

En cas d'erreur, les superviseurs effectuent la correction des données et si besoin est, un retour sur terrain doit être fait par le superviseur ou l'enquêteur.

Etape G

Les données corrigées sont 'rectifiées' à la saisie et repassent à *l'étape D*. etc.....**Le nombre de boucle effectué par une vague correspond au nombre de passage.**

Etape H

Les vagues 'propres' sont stockées dans un répertoire créé à cet effet. Le tableau III fournit les statistiques sur le nombre de vague et de passage pour les Enquêtes 1-2-3 et le Recensement Industriel.

Tableau 1
Durée des étapes de l'informatisation

	Phase I	Phase II	Phase III	Recensement industriel
A-B	7 questionnaires par jour	4 questionnaire par jour	200 questionnaires par 2 semaines	(**)
B-C	4 questionnaire par jour	4 questionnaire par jour	200 questionnaires par 2 semaines	(**)
D-E-F-G	50 questionnaires par jour	2 vagues par semaine	2 vagues par semaines	(**)

N.B : ²Les chiffres sont approximatifs et dépendent fortement des opérations de terrain

Tableau 2
Quelques statistiques globales sur la saisie

	Phase I	Phase II	Phase III	Recensement industriel
--	---------	----------	-----------	------------------------

²(**) Les différentes étapes ont été effectuées suivant l'arrivée des questionnaires. Le retour des questionnaires a duré 5 mois et la saisie a eu lieu au fur et à mesure de l'arrivée de ceux-ci.

Nombre d'opérateurs de saisie	15	15	15	15
Nombre de moniteurs de saisie	5	5	5	5
Nombre de superviseurs de saisie	3	3	3	5
Nombre de questionnaires retournés	2500	1134	400	642
Nombre de pages de questionnaires	4 pages pour la fiche ménage et 8 pages pour les fiches individuelles	11 pages	30 pages	12 pages

III.- STRUCTURE DE L'EQUIPE DE SAISIE

L'équipe de saisie est structurée de manière pyramidale, chaque étage du pyramide jouant un rôle particulier:

Le responsable de saisie

Il coordonne l'ensemble des opérations relatives à la saisie, de la conception du masque à la sortie de l'output final.

Les superviseurs de saisie

Leur tâche consiste à gérer le stock de questionnaires retournés et à faire l'envoi des questionnaires à la saisie. Ils doivent aussi jouer le rôle contrôleur de saisie.

Les moniteurs de saisie

Ils sont chargés de dispatcher les questionnaires aux opérateurs de saisie ainsi que d'organiser le système de la double saisie.

Les opérateurs de saisie

La saisie et la vérification (deuxième saisie) des questionnaires leur sont attribuées.

Notons que les masques de saisie ont été développés à la DESAI et que les test de cohérence ont été développés par des consultants externes au projet.

IV-LES OUTPUTS DU MODELE

Structure des résultats

Selon les besoins de l'utilisateur, **la base de données finale a été subdivisée en différents module qui correspondent en général aux modules des questionnaires.** Le format des fichiers de sortie est le format DBASE. Certaines agrégations ont été effectuées

afin d'éviter une trop grande multiplication de fichiers. **Chaque fichier comporte un identifiant qui permet de faire la liaison avec un autre fichier.**

Tableau 3
Statistiques sur les questionnaires et les fichiers

	Nombre de fichiers en sortie	Nombre de vague	Nombre de passage
Enquête Phase I	7 Fichiers	(*)	(*)
Enquête PhaseII	19 fichiers	5	7
Enquête Phase III	15 Fichiers	8	6
Recensement Industriel	20 Fichiers	7	5

(*) Pour la phase I les questionnaires n'ont pas été regroupés par vague

Stockage des résultats

Les fichiers de résultats ont été stockés dans les répertoires dans la machine pentium utilisée pour les observatoires. Le sous-répertoires utilisés sont les suivants:

Enquête Phase 1 : c:\BASE_DON\1995\PHASE1
Enquête Phase 2 : c:\BASE_DON\1995\PHASE2
Enquête Phase 3 : c:\BASE_DON\1995\PHASE3
Recensement industriel : c:\BASE_DON\1995\RECENS

Tout apurement de fichiers par les utilisateurs doit être signalé au responsable de chaque enquête.

V.- LES AMELIORATIONS POUR LES ENQUETES A VENIR

L'idée de base consiste à tirer les leçons du passé.

Optimisation du temps passé pour chaque étape

Le chevauchement est possible pour chaque étape de la modélisation. Bien qu'il y ait eu chevauchement des étapes lors des différentes enquêtes déjà effectuées, la gestion optimale n'a pas pu être atteinte. **L'éclatement des données et les tests de cohérence peuvent être effectués dès que les données soient saisies.** Il n'est point besoin d'attendre que tous les questionnaires soient saisis avant d'entamer la première étape. Le tout est un problème d'organisation. Il est indispensable que le responsable de saisie consacre une grande partie de son temps au suivi des différentes phases de l'informatisation.

Meilleure collaboration entre le responsable de l'enquête et le responsable informatique

Le responsable de saisie doit informer le responsable de l'enquête des états d'avancements des travaux. Ce dernier devra de son côté se mettre au courant des différents problèmes qui se posent au niveau de la saisie.

Développement du masque de saisie

- Dès que le questionnaire final est arrêté :
- le développement du masque de saisie doit être entamé. Un test sur un questionnaire réel devrait s'ensuivre.
 - les test de cohérence devront être formulés et appliqués au cas réel des questionnaires pilotes.

VI.- LES COUTS FINANCIERS DE L'INFORMATISATION

Les coûts ont été adaptés selon les difficultés rencontrées pour la saisie des données. Les questionnaires de certaines enquêtes étant beaucoup plus difficile à saisir que d'autres. (voir en annexe 7 les coûts unitaires).

VII.- A PROPOS DES ANNEXES

Les annexes contiennent les différents dossiers de base qui découlent des enquêtes. Leur lecture est indispensable dans la mesure où ils constituent les « dessous des enquêtes ».

ANNEXE 1 Les tests de cohérence.

Ils sont formulés par le responsable de l'enquête et les superviseurs. **Les contrôles de champ ont été appliqués systématiquement pour tous les modules.**

ANNEXE 2 Fiche signalétique des logiciels IMPS PERFORM PRO

ANNEXE 3 Fiche de remise de questionnaire

Cette fiche est à remplir systématiquement par les superviseurs d'enquête

ANNEXE 4 Un exemple de circulaire

Il sert à transmettre les messages au différents superviseurs. Ce mode de diffusion des informations est très important car il permet de mettre à un même niveau d'égalité tous les superviseurs.

ANNEXE 5 Fiche de gestion des vagues de questionnaire

ANNEXE 6 Fiche de gestion des micro-ordinateurs

Cette fiche permet de voir les dernières mises à jour des vagues de questionnaires. Elle se remplit verticalement.

ANNEXE 7 Les coûts des enquêtes.

Il s'agit des coûts unitaires par opérateur de saisie, moniteur de saisie et superviseur de saisie; leurs effectifs respectifs étant déjà énumérés dans le tableau II.

ANNEXE 8 Les machines utilisées et leur disponibilité

Un tableau récapitulatif donnant les caractéristiques des machines utilisées et leurs disponibilités. Ce tableau sert à gérer le stock de questionnaires retournés et qui devront être envoyés à la saisie.

Ils serviront de base de calcul pour les enquêtes futures

ANNEXE 1

Les tests de cohérence

Les contrôles de champ ont été appliqués systématiquement pour tous les modules de tous les questionnaires.

PHASE I

Les programmes informatiques sous SAS de ces tests existent au projet MADIO. Les détails des test de cohérence peuvent y être tirés.

MODULE CONCERNE	LES TEST DE COHERENCE
Activités principale et secondaire	Expérience supérieure à l'âge
Activités principale et secondaire	Fonctionnaire non salariés
Activités principale et secondaire	Patron ou à son propre compte et public
Activités principale et secondaire	Expérience supérieure à la durée de l'emploi
Activités principale et secondaire	Ancienneté supérieure à la durée de l'emploi
Activités principale et secondaire	Personne seule non à son propre compte
Activités principale et secondaire	Ancienneté = durée => pas de promotion
Activités principale et secondaire	Auto création d'emploi et non patron ou à son propre compte
Activités principale et secondaire	Tenue de comptabilité et non patron ou à son propre compte
Activités principale et secondaire	Bulletin de paye et patron ou à son propre compte
Activités principale et secondaire	Salaire fixe et non salarié
Activités principale et secondaire	Bénéfice et non patron ou à son propre compte
Activités principale et secondaire	Patron ou à son propre compte et revenu différents de bénéfices
Activités principale et secondaire	Appartenance à un syndicat et pas de syndicat
Activités principale et secondaire	Prestation CNAPS et entreprise non enregistrée à la CNAPS
Activités principale et secondaire	Nombre d'année d'expérience supérieure à emploi actuel et emploi antérieur
Activités principale et secondaire	Durée emploi antérieur supérieur à l'âge
Chômage	Durée chômage supérieure âge
Chômage	Nombre d'année sans emploi dans la vie d'un homme
Chômage	Année sans emploi dans la vie d'un homme or peu d'études
Chômage	Expérience supérieure à âge
Fiche individuelle	Erreur sur le temps passé à Tana alors qu'il s sont nés ailleurs
Fiche individuelle	Erreur sur le nombre d'année depuis la migration (supérieur à l'âge

Fiche individuelle	Le nombre d'année d'étude ne correspond pas au cycle achevé
Fiche individuelle	Ménage sans ou avec plusieurs chefs de ménage
Fiche individuelle	Des " enfants parents" : - âge < 15 ans et déjà parents ou - (âge parents-âge enfant) < 15 ans

PHASE II

RECENSEMENT INDUSTRIEL

Les programmes informatiques sous SAS de ces tests existent au projet MADIO. Les détails des test de cohérence peuvent y être tirés. Voir les questionnaires pour la signification des modules.

MODULE CONCERNE	LES TESTS DE COHERENCE
IE et compte de résultats	IE24=R1+R2
IE et MO	IE23=Totaux des EFF dans MO10
R et VR	R1a=VR1; R1b=VR3; R1c=VR4
R et VR et MP	R2a=VR2 ; R2b=MP2
ST et R	R1d=ST1b-ST1a R2c=ST2b-ST2a R3b=St3b-ST3a
R et MP	R3a=MP1 R3c1=MP3 R3c2=MP4
R et TS	R5a=TS4 R5b=TS1 + TS2 + TS3
R et MO	R5c1=totaux FP1 R5c2=totaux FP2
R et F	R7a=F1 R7b=F2 R9a=F3
MP	La somme des MP14 doit être inférieur ou égale à MP1
VR	La somme des VR13 doit être inférieur ou égal à VR1
I	IN5a+IN5b+IN5c+IN5d doit être égal à 100 %
ZF	ZF10_1a + ZF10_1b + ZF10_1c + ZF10_2a + ZF10_2b doit être égal à 100 %
R et ST	Voir si les valeurs des stocks dans chacun des rubriques sont estimés en LIFO ou FIFO ou en coût moyen pondéré

ANNEXE 2

Fiche signalétique des logiciels IMPS et PERFORM PRO

Fiche signalétique du logiciel IMPS

Nom

Integrate Microsoft Processing System

Modules

IMPS est composé de Huit (8) modules (voir tableau):

Module	Attribution
CENTRACK	Gestion de document
CENTRY	Saisie et Ecran de Saisie
CONCOR	Redressement des données : selon les modalités de redressement préconisées par les analystes
QUICKTAB	Tabulation rapide : sortie rapide de résultat
CENTS	Programme de tabulation : sortie de résultat avec mise en forme des tableaux selon le programme élaboré par l'utilisateur
CENVAR	Calcul de variance : résultat obtenu selon le programme de l'utilisateur
TRS	Publication des résultats

Objectif

La prétention de IMPS est d'être un logiciel à " chaîne d'exploitation complète " c'est à dire un logiciel qui traite à la fois les problèmes en amont et en aval (du masque de saisie à la sortie des tableaux)

Expérience

IMPS a été déjà utilisés utilisé pour traiter des dizaines d'enquête dont le recensement en Inde

Avantage et inconvénient

<i>Avantage</i>	<i>Inconvénient</i>
Permet de traiter un nombre très élevé de d'enregistrement. Exemple : recensement en Inde.	Selon les responsables du recensement qui ont déjà beaucoup utilisé ce logiciel, la manipulation de IMPS est assez difficile. Des contacts permanents ont lieu avec les concepteurs au Bureau of Census aux USA.

<p>Permet de sortir des statistiques sur l'état d'avancement de la saisie :</p> <ul style="list-style-type: none"> * comptage de nombre de questionnaire saisi * comptage de nombre de caractères saisis * calcul automatique de la performance des opérateurs de saisie 	<p>Selon les utilisateurs de IMPS, l'utilisation de ce logiciel n'est pas préconisée si le nombre d'enregistrement est inférieur à 1 million.</p>
<p>Permet de faire la saisie contrôle : Il est possible de donner des instructions à IMPS pour qu'il fasse les contrôle des données saisies</p>	<p>Malgré la "chaîne d'exploitation complète" qu'il propose, IMPS n'est intéressant que pour effectuer le masque de saisie. Les statistiques sorties par IMPS se limitent à un niveau relativement modeste.</p>
<p>Les données obtenues après la saisie sont sous format " ASCII " donc récupérable par facilement par d'autres logiciels.</p>	
<p>Permet d'effectuer la double saisie : Il s'agit là d'un " saisi vérif ". Un questionnaire déjà saisi est resaisi par un autre opérateur ; si celui-ci entre une valeur différente de celle qui est à l'intérieur de la base de données alors un " bip " sonore intervient.</p>	

Fiche signalétique du logiciel Perform Pro

Modules

Perform Pro est composé de Deux (2) modules :

Module	Attribution
COCEPTION	Conception de questionnaire
REPLISSAGE	Remplissage de questionnaire

Objectif

Perform Pro permet de concevoir, de gérer, de remplir et d'imprimer des formulaires.

Avantage et inconvénient

<i>Avantage</i>	<i>Inconvénient</i>

<p>Perform Pro est un logiciel très convivial : la forme de l'écran de saisie est exactement celle du questionnaire.</p>	
<p>Permet de faire la saisie contrôle : Il est possible de donner des instructions à Perform Pro pour qu'il fasse les contrôle des données saisies</p>	<p>Ne permet d'effectuer la " saisie vérif " (voir fiche signalétique de IMPS pour la signification "saisie vérif")</p>
<p>En matières de gestion de données; Perform Pro peut facilement gérer les Tables Multi-Fiches ou les Fiches Multi-Tables ou des Fiches et Tables multiples.</p>	
<p>Les données obtenues après la saisie sont sous format " ASCII " ou DBASE donc récupérables par facilement par d'autres logiciels.</p>	

ANNEXE 3

Fiche de remise de questionnaire

Codification Phase 2			
Superviseur			
Nombre			
Date de livraison			
Code	Rectification	Code	Rectification
1-	1-	1-	1-
2-	2-	2-	2-
3-	3-	3-	3-
4-	4-	4-	4-
5-	5-	5-	5-
6-	6-	6-	6-
7-	7-	7-	7-
8-	8-	8-	8-
9-	9-	9-	9-
10-	10-	10-	10-
11-	11-	11-	11-

ANNEXE 4

Exemple de circulaire

NOTE SUR L'INFORMATIQUE

ECRIRE SYSTEMATIQUEMENT AU " **STYLO**" LA CODIFICATION

FICHE MENAGE

Ne pas remplir la partie " ANARAN'NY MPANAO SAISIE "

QUESTIONNAIRE INDIVIDUEL

- 1 -Ne pas oublier de remplir la partie " MARIKA FANAVAHANA "
- 2 -Remplir avec 2 (deux) chiffres la partie " FOTOANA NIFAMPIRESAHANA "
exemple : ANDRO : 09 VOLANA : 04 TAONA : 95

- 3 -Remplir de la même manière que l'exemple suivant la partie " FAHARETAN'NY FANADIHADIANA "

exemple : Ora nanombohana : 10H30 Ora niafarana 11H45
(ne pas oublier le "H" pour l'heure).

- 4 - Remplacer manuellement :

* par une (1) case les deux cases dans la partie " ANARANN'NY CODIFICATEUR "
et

* par deux (2) cases la partie " ANARAN'NY MAPANO SAISIE "

- 5 - Remplir la partie " ANARAN'NY CODIFICATEUR " par le numéro du superviseur qui est censé lui-même effectuer la codification

- 6 - Pour la question EA1 mettre systématiquement " 00 " (zéro zéro) pour le nombre d'heures au cas où la personne n'est pas concernée par la rubrique proposée (1,2,3,4).

- 7 - Si la réponse est manquante (la personne ne veut pas répondre alors que la case doit être remplie), mettre le signe " = "(égal) dans la case à codification.

- 8 - les parties 6 et 7 servent à faire la distinction entre les valeurs nulles, les blancs, et les réponses manquantes.

- 9 -Cas de la question TP16a :

Si la réponse est " tsy mifidy " codifier cette réponse par "00" (double zéro).

ANNEXE 5

Fiche de gestion de vagues de questionnaires

	VAGUE1	VAGUE2	VAGUE3	VAGUE4	VAGUE5	VAGUE6	VAGUE7	VAGUE8
PASSAGE 1	X	X	X	X	X	X	X	X
PASSAGE 2							X	X
PASSAGE 3								
PASSAGE 4								
PASSAGE 5								
PASSAGE 6								
PASSAGE 7								

ANNEXE 6

Fiche de gestion de micro-ordinateurs

VAGUE 1	VAGUE 2	VAGUE N
Micro1	Micro 2	Micro 3		Micro 4
		Micro 2		Micro 1

ANNEXE 7

Les coûts unitaires

	PHASE I	PHASE II	PHASE III	Recensement Industriel
Opérateur de saisie	170.000 Fmg	170.000 Fmg	170.000 Fmg	200.000 Fmg
Moniteur de saisie				300.000 Fmg
Programmeur	500.000 Fmg	500.000 Fmg	500.000 Fmg	900.000 Fmg
Superviseur de saisie	500.000Fmg	500.000 Fmg	500.000 Fmg	500.000 Fmg

ANNEXE 8

Les machines utilisées et leurs disponibilités

Les machines utilisées	Caractéristiques
Type de machines	IBM PS/1 386SX 25MHZ RAM 2MB
Nombre	14
Propriétaire	INSTAT
Disponibilité	2 semaines ouvrables par mois
Contraintes	Sode des fonctionnaires, Pension, Fiche électorale, imprévus