



STATISTICS
DENMARK



Statistisk sentralbyrå
Statistics Norway



Statistiska centralbyrån
Statistics Sweden

MZ:2012:07p

CSPro e processamento de dados do Inquérito Contínuo multi-objectivos INCAF

Relatório duma missão ao Instituto Nacional de Estatística de Moçambique,
Maputo, Moçambique

27 Agosto - 5 Outubro 2012

dentro do quadro do

**CONTRATO PARA CONSULTORIA
EM CAPACITAÇÃO INSTITUCIONAL,
ESTATÍSTICAS ECONÓMICAS E ÁREAS AFINS**

entre

INE e Scanstat

Julio Ortúzar



INSTITUTO NACIONAL DE ESTATÍSTICA

SERPRO S.A.

Julio Ortúzar

joc@serpro.com

Conteúdo

Resumo	3
Summary in English	3
1. Tarefas efectuadas durante o período de 6 semanas	4
1.1 Principal tarefa: organização e ordem para o trabalho de campo (captura de dados).....	4
1.2 Usando CAPI para melhorar a qualidade dos dados	5
1.3 Reestruturação dos dados existentes, programas de captura.....	7
1.4 Desenvolvimento do módulo INFOR	9
2. FUTURAS ATIVIDADES DE TI.....	11
2.1 Melhorias para a implementação do sistema anterior.....	11
2.2 Manutenção do sistema atual	11
2.3 Preparação do layout ou formato exclusivo para as diferentes ondas.....	11
2.4 A produção das tabelas para análise.....	12
2,5 Divulgação de dados de la Pesquisa.....	12
3. OBSERVAÇÕES FINAIS.....	12
3.1 O teste de sistema	13
3.2 Ficheiro da Mostra	13
3.3 Arquivo de dados equipes	14
Anexo 1 ToR	15
Anexo 2 Sistema de controle CAPI.....	18
1. Introdução.....	18
2. Instalação de máquinas.....	19
2.1 Identificação da máquina.....	20
2.2 Definição do número IP (Windows 7)	23
2.3 Estruturas de pasta e arquivos a serem transferidos	25
3. Alguns arquivos de dados importantes que precisam ser gerados.....	29
3.1 Arquivo de dados equipes.....	29
3.2 Exemplo de arquivo de dados	31
3.3 O arquivo de controle	31
4. Sistema de controle a orientado por menus.....	33
4.2 Menu a autoridades de supervisão	36

Resumo

Durante o período de seis semanas no INE-Moçambique, foram realizadas as seguintes tarefas:

- (1) A completa reestruturação do antigo sistema INCAF, proporcionando mais liberdade para as entrevistas pessoais no sentido que membros do agregado familiar poderiam ser seleccionados para ser entrevistado de acordo com as necessidades dos entrevistadores.
- (2) A execução do sistema de controle de operação para adicionar uma administração sistemática de entrevistas e assim facilitar o fluxo de dados dos entrevistadores para os supervisores e para o servidor central.
- (3) Uso de CAPI para melhorar a qualidade dos dados (implementação da codificação assistida e verificação para o módulo INFOR (INFORmal sector económico da população) de consistência).
- (4) Desenvolvimento e implementação do módulo INFOR.

Tarefas restantes para actividades futuras são:

- (1) melhorias para a implementação do sistema anterior.
- (2) Periódica manutenção do sistema actual.
- (3) Elaboração do formato comum e sistema de codificação para as diferentes ondas da pesquisa.
- (4) A produção de tabelas para análise e controle do trabalho de campo.
- (5) Divulgação de dados usando a WEB como a maneira mais conveniente para difundir as tabulações e outros documentos.

Summary in English

During the six week period at INE-Mozambique, the following tasks were performed:

- (1) A complete re-structuring of the old INCAF system, providing more freedom to the personal interviews in the sense household members could be selected to be interviewed according to the interviewers' needs.
- (2) Implementation of the Operation Control System to add a systematic administration of the interviews and facilitating the flow of data from the interviewers to the supervisors and from them to the central server.
- (3) Using CAPI to improve the quality of data (implementation of the assisted coding and consistency checking to the INFOR module (INFORmal economic sector of the population).
- (4) Development and implementation of the INFOR module.

Tasks remaining for future activities are:

- (1) Improvements to the previous system implementation.
- (2) Periodic maintenance of the current system.
- (3) Preparation of the common format and coding system for the different waves of the survey.
- (4) The production of tables for analysis and control of the fieldwork.
- (5) Data dissemination using the web as the most convenient way to spread out the tabulations and other documents.

1. Tarefas efectuadas durante o período de 6 semanas

O principal objetivo desta missão foi ajudar ao INE na implementação de um sistema contínuo de múltiplos propósitos, mais especificamente um sistema de captura de dados. Um objetivo secundário, mas não menos importante foi/é ajudar ao INE no design da base para a concepção de um sistema de processamento completo para esta importante pesquisa, incluindo a consolidação de dados, conversão de dados em um formato único para todas as ondas diferentes, tabulação de dados e disseminação de dados.

Durante os primeiros dias da missão, foi efectuada uma avaliação rápida da situação de processamento de dados. Os resultados não foram promissórios, uma vez que foi evidente que: (1) a captura de dados implementada para a primeira onda foi simplesmente uma coleção de programas sem qualquer ligação entre eles; (2) os programas foram questionários eletrônicos sem qualquer verificação de consistência, capturando os dados de cada questionário: habitação (família adicionada), o questionário para capturar dados de todos os membros do agregado familiar de 5 anos de idade e mais velhos e o questionário para capturar os dados de todos os membros do agregado familiar de 5 a 17 anos de idade trabalhando; (3) os últimos dois questionários, sendo dos mesmos membros domésticos enumerados e completamente identificado por questionários previos deveria definir quem deveria e quem não deveria ser parte de cada um deles; No entanto, não tendo qualquer ligação entre eles, esta importante tarefa não foi verificada.

A primeira avaliação deu cerca de 2,5% das pessoas de cinco anos e mais velho (segundo questionário) e 17% dos questionários terceiros (5-17 anos de trabalho de idade) em falta.

Outra questão importante foi também evidente: houve falta de organização na operação de campo.

Após a primeira avaliação, uma reunião com o diretor, Arao Balate, foi para explicar a situação atual e propor meu plano de trabalho para superar, pelo menos, as deficiências mais importantes INCAF nos dados resultantes do processamento.

Durante a reunião, foram explicados os principais problemas resultantes do processamento de dados e apresentou-se o seguinte plano:

1.1 Principal tarefa: organização e ordem para o trabalho de campo (captura de dados)

A tarefa mais importante e urgente para a realização foi organizar e ordenar a captura de dados e trabalho de campo que naquele momento estava bastante caótica. Para atingir este objetivo, ofereci o sistema de controle de operação projetado e desenvolvido pelo SERPRO. Este sistema foi implementado em vários países - muitos na África - e haveria um avance muito importante na economia de tempo e

recursos se pudessemos implementá-lo para o projeto INCAF. O sistema de controle exigiria uma fase de adaptação importante, pois o inquérito já estava em andamento, mas os resultados devem compensar o esforço. Houve no entanto alguns requisitos básicos que seria preciso cumprir a fim de ter sucesso com a adaptação:

- Seria necessário um arquivo de dados de amostra com todas as informações requeridas pelas metas muito exigentes da pesquisa painel INCAF. Desde que era uma pesquisa contínua, não só se precisava da localização geográfica dos domicílios a serem entrevistados mas também seriam fornecidos os endereços de cada agregado familiar juntamente com os nomes e características demográficas básicas dos seus membros do agregado. No final, eu decidi usar o arquivo de dados domésticos produzido durante a primeira rodada para gerar o arquivo de dados de controle. Durante este esforço, várias deficiências dos dados foram reveladas, mas finalmente, conseguimos criar um arquivo parcialmente concluído. Parcialmente porque a primeira onda estava ainda em curso e, portanto, o arquivo de controle final deveria aguardar a conclusão desta ronda.
- Informações detalhadas sobre a composição de cada grupo (nomes dos supervisores com identificação dos seus respectivos entrevistadores), bem como os clusters atribuídos a cada um dos supervisores. Esta informação é crucial para criar o ambiente correto para o sistema de controle operar com êxito.
- O controle do sistema tinha sido projetado para operar utilizando o FTP (File Transfer Protocol) para transferir dados e software desde a cabeça, o diretor de TI responsável pela INCAF projeto, até os entrevistadores e vice-versa. Assim, os supervisores e entrevistadores necessitavam uma conexão de rede simples para ser capazes de usar os nossos utilitários de conexão (servidor e cliente). A maneira mais simples e fácil foi a conexão usando um cabo de par trançado através da interface RJ-45. Desde que os tablets em uso por entrevistadores não tinham o RJ-45 para ligar para os netbooks usados pelos supervisores, INE concordou em obter um adaptador para cada supervisor. Haveria muito tempo para obter os adaptadores de rede e, nesse ponto, não houve nenhum problema previsto.
- Uma explicação completa e documentação do sistema de controle foi dado ao INE. A documentação fornecida ao guia para gerenciar todo o sistema é suficientemente simples para ser entendido por qualquer pessoa de TI, mas principalmente pelo diretor de projeto de TI. (Anexo 1)

1.2 Usando CAPI para melhorar a qualidade dos dados

Como foi dito, a implementação atual foi apenas um questionário eletrônico e, portanto, não houve nenhum ganho na entrevista tipo CAPI. Para aproveitar ao máximo as oportunidades que o CAPI oferece para melhorar a qualidade dos dados recolhidos, havia duas tarefas que podiam ser incorporadas aos programas atuais: a codificação assistida variável de tipo aberto que neste projeto foram muitas, e a verificação durante a entrevista de consistência. Dessas duas tarefas, apenas a primeira pudo ser

implementada durante a minha primeira visita, devido as restrições de tempo. A segunda teve que ser deixada pendente para o próximo período. SERPRO é o proprietário do software CSProX, uma versão estendida do CSPro, que é capaz de executar a codificação assistida (entre muitas outras extensões que o sistema público não tem). Assim, também foi acordado que CSProX seria usado para tirar proveito do sistema CAPI (sem custo para o INE).

O algoritmo de codificação assistido criado, projetado e realizado pelo SERPRO é uma metodologia poderosa que permite chegar ao código adequado de duração indeterminada, variável de forma simples e rápida como segue:

- O usuário, neste caso o entrevistador, insere uma ou mais palavras-chave que descrevem melhor a descrição da variável sendo codificada. O sistema aplica a cada palavra-chave el algoritmo de auto-completar (auto-complete) exibindo do dicionário de termos armazenados para a variável, todos aqueles que têm o mesmo prefixo; como o operador acrescenta mais letras para a palavra-chave, a lista de termos diminui em número até que ele chega a um único termo/palavra-chave. Normalmente duas ou três letras produzem uma listinha dela que, a palavra-chave adequada pode ser selecionada. Se a palavra-chave inserida não está no dicionário de termos (criado para cada variável a ser codificada com base na descrição do código), o termo entrado não tipifica a descrição da variável, e não deve ser usado. Muitas vezes, um sinônimo pode ser usado (ou seja, professor ou maestro, etc.). Depois que a primeira palavra-chave tem sido inserida, existem duas opções, a busca de todas as descrições que contêm o termo que acabase de entrar, ou a adição duma ou mais palavras-chave para melhor tipificar a descrição. Em muitos casos, uma palavra-chave produz uma curta lista de descrições suficiente para selecionar o termo desejado; Se a lista for muito grande, o sistema pedirá uma qualificação mais adequada ou melhor da descrição. No nosso caso específico, algumas das listas foram muito complicadas incluindo o nome de produtos quimicos que normalmente requerem dois ou três palavras-chave para produzir uma pequena lista de alternativas.
- O método como se explicou vai analisar as palavras-chave inseridas pelo entrevistador e se elas são vagas enquanto não produzem uma lista selecionável, se exigirá uma melhor descrição. Isto é importante, ja que a alternativa que basta para inserir a descrição a ser codificada mais tarde no escritório central, produzirá inevitavelmente um código que não é exato. Assim, esta abordagem é um passo em frente na tarefa de melhorar a qualidade dos dados recolhidos.
- Além dos benefícios já mencionados, o algoritmo de autocomplete fornece uma alternativa mais rápida para digitar as palavras-chave completas, especialmente utilizando os tablets disponíveis para a realização das entrevistas.
- Por último, há também uma importante poupança em tempo no escritório central uma vez que os dados provenientes do campo estão prontos para análise.

1.3 Reestruturação dos dados existentes, programas de captura

Como foi mencionado anteriormente, os três principais programas para conduzir as entrevistas foram usados independentemente um do outro. Isto implicava que, em muitas ocasiões, uma pessoa, que teve de responder, por exemplo, os três questionários, teve que esperar até o turno adequado para responder o primeiro, em seguida, o segundo e, finalmente, o terceiro questionário. Este fato foi causa de desconforto, fadiga e uma boa razão para rejeitar a longa entrevista.

Uma tarefa importante foi a reestruturação dos programas em dois aplicativos: o módulo doméstico data capture e as entrevistas pessoais, que fundiram numa aplicação os três questionários já mencionados: 5 anos e mais velhos membros do agregado familiar (um entrevista para cada membro), 5-17 anos de idade trabalhando atualmente e o recém INFOR ou módulo para o sector INFORmal da população (que estavam trabalhando para seus próprios membros - com ou sem empregados). A seguinte seqüência de capturas de tela vai ajudar a compreender a importância da reestruturação de tarefas.

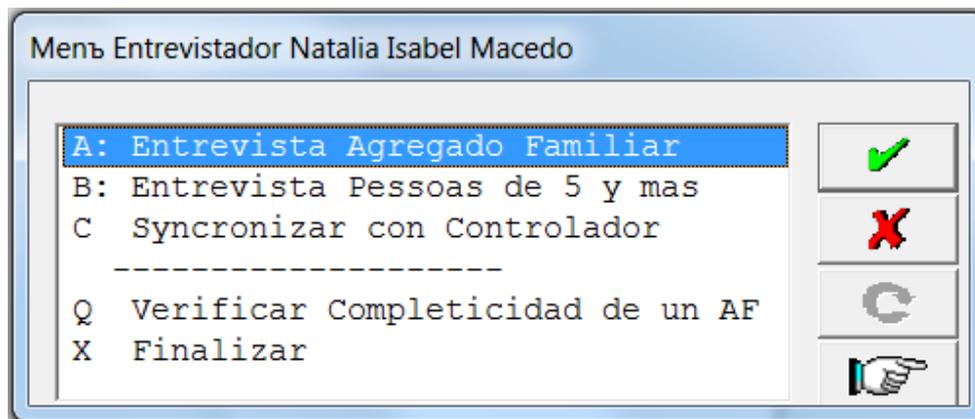


Figura 1: Tela do Menu do entrevistador

As duas primeiras opções oferecem ao entrevistador a realização da entrevista do agregado familiar e a entrevista pessoal (por 5 anos e mais velhos). A terceira opção permite sincronizar com a máquina do supervisor para transferir os dados duma máquina a outra e/ou atualizar o software no caso de ter uma versão atualizada. A quarta opção, muito importante, permite ao entrevistador verificar todas as entrevistas realizadas em uma casa particular e ver se há uma ou mais entrevistas incompletas.

Selecionando a primeira opção, o entrevistador irá obter uma lista das famílias a ser entrevistadas como segue:

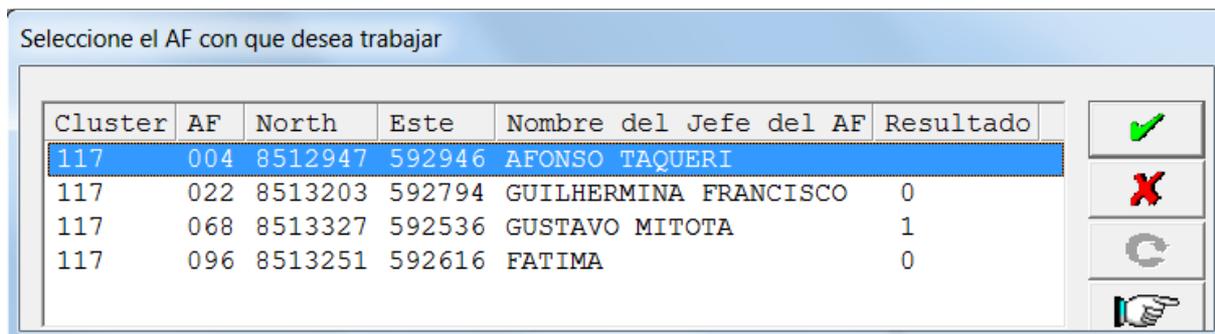


Figura 2: Lista de famílias para a entrevista

O entrevistador tinha que entrevistar quatro famílias identificadas pelo número de cluster, o número de doméstico seguido das coordenadas GPS e o nome do chefe do agregado familiar. No final, o resultado da entrevista também foi exibido para que o entrevistador pudesse ver quanto trabalho faltava (o terceiro membro mostra que a entrevista foi concluída e os outros três não tinham começado ainda).

Ao seleccionar qualquer uma das três famílias com nenhuma entrevista ainda, o sistema iniciará uma nova entrevista, fornecendo todas as informações básicas para esse agregado familiar obtido a partir do arquivo de controle. Por outro lado, se o entrevistador selecciona aquele com a entrevista completa, o sistema emitirá uma mensagem de aviso e recuperará a entrevista terminada para inspecionar ou fazer qualquer outra alteração necessária.

Se o entrevistador selecciona no menu inicial a opção de entrevistar 5 anos e mais velho (entrevista pessoal), o sistema exibirá a mesma tela mostrada na primeira figura 2 para que o entrevistador tenha a oportunidade de definir o doméstico com o que quiser trabalhar e em seguida, a lista de membros de famílias (s) que são 5 anos ou mais velhos.

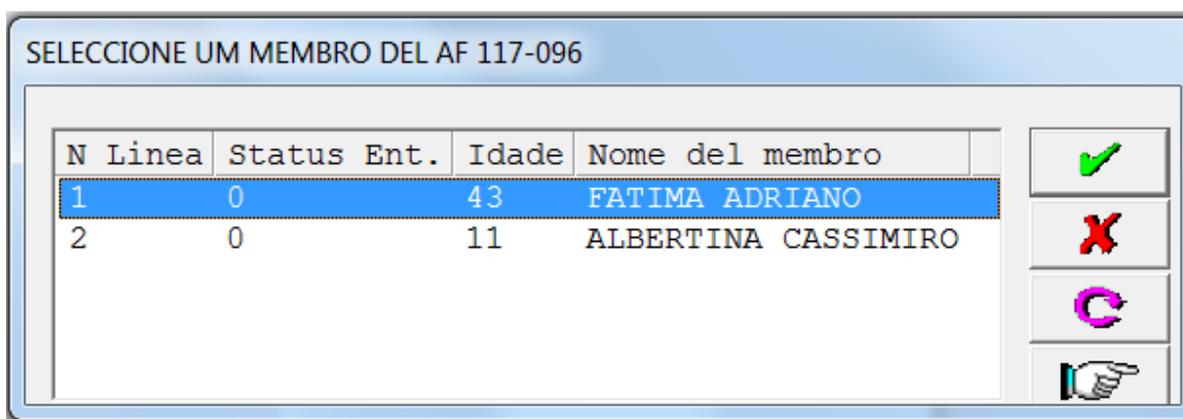


Figura 3: Lista de membros do agregado familiar para a entrevista

A tela mostrada como figura 3 é o correspondente para seleccionar a casa de Fátima (última mostrada na Figura 2) e corresponde a todos os membros do agregado familiar de mais de 4 anos de idade.

É importante apontar que aqui o entrevistador fica livre para selecionar qualquer um dos membros da lista, desconsiderando a ordem na qual eles são listados. Esta é certamente uma grande vantagem, desde que o membro que está atualmente disponível para a entrevista pode ser iniciado primeiro. O membro selecionado passará por diferentes questionários em um fluxo contínuo e, se não qualifica para um questionário específico, ele será simplesmente ignorado.

Novamente, o resultado da entrevista também é exibido (Status Ent.) para que o entrevistador possa estar ciente do status do trabalho.

Finalmente, o entrevistador tem a opção de verificar a conformidade de entrevistas para uma determinada família. O agregado é selecionado a partir do mesmo ecrã mostrado na Figura 2 e o sistema, depois de verificar o Status ou os resultados de várias entrevistas pessoais, exibirá a seguinte tela:

N Linea	Questionario	Status Entrevista	Idade	Nome del miembro
1	5 y Mas	Ainda nao empezado	57	GUSTAVO MITOTA
2	5 y Mas	Ainda nao empezado	38	FILOMENA MASSALO
3	5 y Mas	Ainda nao empezado	8	TINO FERNANDO
4	5 y Mas	Ainda nao empezado	11	JULIETA FERNABDO

Figura 4: Completude das entrevistas individuais

Esta longa explicação é dada porque mostra claramente a importância da reestruturação realizada e concentra o esforço em fornecer as ferramentas para realizar as entrevistas de forma simples e eficiente. Este esquema pressupõe que as informações fornecidas pelo INE em relação à composição do agregado familiar e os endereços estão completos e corretos, e que as informações das equipes com a atribuição de cluster são confiáveis. Esses dois arquivos são a base sobre a qual o sistema foi concebido e implementado.

1.4 Desenvolvimento do módulo INFOR

Este módulo tem grande importância tanto para o INE como para o governo, já que concentra a atenção no estudo dos fatores sócio-econômicos das pessoas que pertencem ao sector económico informal(mais instável) que necessitam de atenção especial do governo.

O módulo é bastante grande e complexo; Tem 5 telas de entrada de dados para o módulo agrícola e 19 para o módulo não-agrícola, ou seja, aproximadamente 30% da pesquisa toda.

Um dos principais problemas encontrados no desenvolvimento do software CAPI para este módulo foi que, no momento, o questionário estava ainda na fase de melhorias e mudanças. Isto implicava que, em

algumas ocasiões, a abordagem para implementar o aplicativo teve que ser drasticamente alterado. É importante apontar que as últimas alterações ao questionário foram feitas em outubro 10,th, e o novo questionário foi enviado para mim em outubro 24th. Eu só ouvi falar sobre a existência deste novo questionário falando com o projeto-chefe sobre alguns problemas com o módulo INFOR. Ele estava se referindo a estas mudanças mas nunca enviou o novo questionário para mim.

Bem, o desenvolvimento deste módulo foi doloroso devido à falta de ajuda de meu colega. Deveria ter sido mais fácil se eu tivesse contado com a contribuição necessária, particularmente o sistema de teste.

2. FUTURAS ATIVIDADES DE TI

2.1 Melhorias para a implementação do sistema anterior

Como se viu nos números anteriores, o trabalho já feito é muito grande, considerando o intervalo de tempo em que foi realizado. No entanto, a melhoria do sistema anterior ainda precisa ser feito em termos de adição a necessária consistência lógica de verificação. Pelo menos uma semana das estimadas três semanas para adicionar o controle necessário tem sido utilizada no ajuste final do sistema atual. Em termos de tempo total alocado para esta tarefa, pelo menos 4 semanas devem ser orçadas.

2.2 Manutenção do sistema atual

INCAF é um projeto grande e ambicioso e além disso, ele altera de onda para onda adicionando novos módulos para os já existentes e/ou suprimindo módulos previos. Portanto, o sistema está constantemente mudando, exigindo uma atividade de manutenção constante.

Embora a amostra sobre a qual a pesquisa é realizada não deve mudar, algumas famílias novas podem assumir o lugar de antigas famílias que mudaram-se para fora. Assim, há um problema futuro em termos de análise e comparabilidade de ondas diferentes, a menos que algumas ações sejam projetadas e implementadas.

Eu poderia alocar mais 4 semanas de tempo (1) estritamente manutenção; (2) a evolução para a terceira onda. Mais para fazer uma estimativa justa, seria necessário ver o módulo questionário.

2.3 Preparação do layout ou formato exclusivo para as diferentes ondas

A fim de preparar as informações de alcance coletadas, é necessário projetar o denominador comum dos arquivos de dados recolhidos em cada fase. Isto implica que, fisicamente falando, o mesmo dicionário de dados funcionará para qualquer um dos vários arquivo de dados de onda.

Isto ajudará a ter apenas um aplicativo de exportação para os três principais pacotes estatísticos (SPSS, Stata e SAS) e para qualquer arquivo de dados de onda. Além disso, o arquivo de dados resultante em qualquer onda deve ser mesclado em um arquivo de dados final tendo como identificador o número de onda. Isto também facilitará a análise comparativa de tempo entre as ondas diferentes.

Eu poderia alocar um tempo de 1 semana para essa tarefa.

2.4 A produção das tabelas para análise

Um tempo deve ser alocado para essa tarefa, mas ao não ter a especificação deles, é difícil fazer qualquer estimativa. É evidente que a julgar pela quantidade de informações recolhidas, o número de tabelas será enorme.

Minha idéia seria criar um sistema integrado que pudesse ser facilmente manipulado por um não-especialista em TI, por menus, que poderiam cuidar da integração de dados, exportação de dados e tabulação de dados. Algumas das primeiras tabelas devem ser orientadas para fornecer indicadores sobre a produtividade das equipes diferentes, integralidade das entrevistas, certificar-se de que não há nenhum resultado de entrevista "em andamento", etc. Essas tabelas de fato devem ser projetadas e implementadas, o mais pronto possível. Portanto, a onda atual pode estar sob constante vigilância em breve. Já que existem dois tipos de questionários, a tabela deve contemplar diferentes unidades pela equipe e pela produtividade do entrevistador: número de agregados familiares entrevistados e número de membros do agregado familiar 5 e mais velhos.

2,5 Divulgação de dados de la Pesquisa

Se eu entendi corretamente, a intenção do INE é criar um sistema de disseminação usando a web para divulgação dos dados de la inquerita. Obviamente este esforço poderia ser compartilhado com todos os outros projectos ou estudos realizados por Instituto Nacional de Estatística.

Uma vez mais, aqui temos uma diferença muito grande no custo dependendo do que queremos fazer. Provavelmente, a idéia mais conservadora seria simplesmente publicar as tabelas ou quadros mais importantes classificados por assunto. Talvez, poderíamos ter um pequeno motor de busca por diferentes keys (ou seja, assunto - educação, atividade, ano, etc-) e o motor de busca seria fornecer uma lista das informações disponíveis para cada conjunto de parâmetros.

Este projeto poderia ser de 3 ou 4 anos que poderiam começar a fornecer o mínimo exigido e adicionando mais complexidade e poder a cada ano.

3. OBSERVAÇÕES FINAIS

Eu acho que é importante ressaltar que a tarefa realizada foi demasiado grande para ser realizada no tempo estimado. Havia muitas pontas soltas e talvez, tenha sido um equívoco da minha parte pensar que a tarefa empreendida seria menos demorada. O fato é que ela exigiu um esforço incrível de minha parte para ter o sistema funcionando como agora.

Talvez eu estava esperando uma maior ajuda do meu colega do que o que eu teve. Ao longo de minha vasta experiência, aprendi que não é bom nem construtivo expressar uma opinião negativa sobre a contraparte, mas neste caso em particular, sinto-me obrigado a fazê-lo pois acho que ele não é a pessoa adequada para dirigir as atividades de TI para este importante projeto. Minha opinião é que também

não queria cooperar ou que ele foi incapaz de fazê-lo. O simples fato é que algumas tarefas que lhe pedi fazer especificamente, não foram feitas ou simplesmente não foram feitas corretamente.

Os principais problemas encontrados que tornaram meu trabalho mas difícil podem ser sintetizadas da seguinte forma:

3.1 O teste de sistema

Uma tarefa crucial que precisava ser feita era um teste exaustivo de todo o sistema. Em uma pesquisa CAPI, diferentes questões precisam ser cuidadosamente testadas: a lógica e particularmente o padrão ignora, as mensagens de erro (ainda existem as mensagens de erro em espanhol, uma vez que sou incapaz de escrever Português e as perguntas que precisam ser exibidas para os entrevistadores (o mesmo de antes), etc.) O fato é que apenas voltei ao meu escritório teve o feedback dos diferentes testes realizados por ele ou sua equipe.

Questões como a codificação assistida precisavam de atenção especial em testes, dada a grande responsabilidade que é implícita. O fato é que o texto descritivo das diferentes variáveis para código estava em Português e os caracteres especiais como acentos e inflexões não tinham uma representação ASCII. Assim, peço-lhe que os revise para que esses caracteres especiais sejam suprimidos. Apenas quando eu estava aqui no meu escritório, só um arquivo de três requeridos foi enviado.

3.2 Ficheiro da Mostra

Como foi explicado ao longo deste documento, o arquivo de dados de amostra, es um ficheiro crucial para ter tudo o sistema trabalhando harmónicamente, foi das principal preocupações durante minha estadia em Maputo. Este arquivo foi gerado usando como entrada os dados de Agregados Familiares das entrevistas feitas durante a primeira onda da pesquisa. No entanto, desde o início, houve problemas por causa de chaves duplicadas ou por qualquer outro motivo. Finalmente, antes da minha partida, tivemos um arquivo de trabalho que aparentemente satisfazia as necessidades. Este arquivo não foi completado porque o trabalho de campo estava ainda em curso, mas pelo menos, a parte restante deveria ser integrada ao arquivo principal e mandada para mim para eu puder produzir o arquivo de dados de controle final. Isso foi feito como planejado, e finalmente, não há muito tempo, pudimo ter esse arquivo importante incorporado ao sistema.

Para minha surpresa, durante a primeira semana de trabalho de campo da segunda onda, um dos supervisore me informou que um cluster inteiro tinha membros dos domicílios errados e características; em palavras simples, todas as famílias desse cluster tinham o número errado de cluster. Este poderia ser um problema grave, ou pode ser apenas um problema isolado.

Esperamos que outra experiência como essa não aconteça novamente. O resultado deste acidente é que, uma vez que eu tinha as informações corretas para o cluster, eu tinha que corrigir manualmente o arquivo de controle do entrevistador e, em seguida, o arquivo de controle do supervisor. Esta é uma

tarefa muito delicada, uma vez que o arquivo de controle armazena todo o trabalho histórico realizado pelos diferentes jogadores deste sistema. Se o mesmo acontecer mais tarde, o trabalho será muito mais difícil, uma vez que mais informações serão adicionadas. Vamos esperar que não encontremos qualquer outro problema similar.

3.3 Arquivo de dados equipes

Como apontado anteriormente, as arquivo de equipes lojam toda a composição de equipe fornecendo informações muito importantes sobre as UPAs atribuídos para cada supervisor, a identificação da máquina e o código para cada entrevistador e supervisor juntamente com seus respectivos nomes. Este arquivo é usado cada vez que há sincronização entre dois computadores e as informações fornecidas permitam que o sistema aceite ou rejeite o processo de sincronização. Além disso, a atribuição de clusters para cada grupo permite gerar o arquivo de controle.

O arquivo de equipes fornecido para mim tinha vários erros como máquina duplicados identificações de máquina duplicados e clusters não atribuídos. A negligência na geração deste arquivo é simplesmente incrível. Até agora, não recebi o arquivo das equipes. Ele precisa ser corrigido, caso contrário, os clusters ausentes não vai ser entrevistados.

TERMS OF REFERENCE

for a short-term mission on

**CSPro and data processing at
the Continuous Multi-purpose Survey**

June 2012

within the

AGREEMENT ON CONSULTING IN INSTITUTIONAL CAPACITY BUILDING, ECONOMIC
STATISTICS AND RELATED AREAS
between INE and Scanstat.

Consultant: Julio Ortúzar

Counterparts: Arão Balate and the Continuous Multi-propose Survey team at INE

Background

The yearly survey program at INE has since 2000 collected information on employment, time use, household budget, health, victimization, sanitation, water supply, basic infrastructure and buildings etc. with a periodicity of 5 years for each module

Since the beginning of 2000 studies have shown the importance of poverty assessments as instruments for monitoring development. The lack of timely information on poverty, and wellbeing, is at the moment discussed among users where the frequency and timeliness of household surveys within the information system for poverty monitoring being an important aspect. Current information is required to monitor and adapt to on-going dynamic changes.

Studies show that the lack of employment is one of the two most important social problems in Mozambique according to the inhabitant's own judgment. This is especially a problem for the households where the head only has short-term employment, paid on a daily/hourly basis. Employment is also a major concern among students 15 years and older.

For these reasons INE has decided to modernize the system by introducing a new Continuous and Multi-Propose Survey (INCAF) having a Core Module based on Labour Force and Expenditure where different modules like Informal sector, Tourism, Education etc. will be introduced quarterly.

Data collection for INCAF will be made using PC-tablets. Initially CSPro will be used as INE already has experience of that software.

Main reasons for the mission

INE has gained a lot of experience carrying out household surveys. However INE still lacks experience from Continuous and Multi-propose Household Surveys and thus requires assistance in order to successfully implement this new chain of surveys taken into consideration the continuous modus operandi needing a streamlined work process avoiding bottlenecks, errors and delays.

Objective

The overall objective of the mission is to assist INE in the implementation of the new Continuous Multi-propose Survey using international experiences regarding data collection and especially CSPro. A second objective is to help INE in creating a layout for a complete processing system.

Activities

The Consultant is responsible for the following activities:

1. Verify, complement and document the data collection system for the survey.
2. Prepare validation specifications for all questionnaires.
3. Develop a validation program from the specifications prepared in step (2)
4. Test and implement the validation system.
5. Assist with training of Central Staff to use the data collection system
6. Propose a layout for a complete processing system with a 3 month repeat time, covering all sub processes and transformations from data collection to dissemination on the Internet.
7. Help INE in setting up a scheme for further assistance regarding the data processing of INCAF.

Expected outputs

A fully working data collection system adapted to INCAF. The consultant will prepare a written documentation of the above activities;

Beneficiaries of the mission

The mission will benefit INE staff, interviewers and the whole National Statistical System of Mozambique including users of statistical information.

Tasks to be done by INE to facilitate the mission

- Elaborate ToR for the training
- Prepare and supply the consultant with necessary documents and information, like the new visual identity of INE
- Supply good communication conditions for the consultant.

Source of Funding

Project: MPD-2008-0006 – Inquérito Sobre Orçamento Familiar – IOF
PAAO12 – 1.4.2 Inquérito Contínuo aos Agregados Familiares

Timing of the mission

This first mission will start on August 16, 2012 and last for 6 weeks. Further missions will be agreed upon during the first mission.

Place

The premises of INE in Maputo with possible allocations to the provinces.

Language

Portuguese, Spanish and English.

Report

The consultant will prepare a short final report to be discussed with INE before ending assignment. Statistics Denmark as Lead Party will publish the final version on www.dst.dk/mozambique within 3+ weeks of the end of the mission. The structure of the report should be according to Scanstat format.

Anexo 2 Sistema de controle CAPI

Sistema de controle CAPI

27 De Julho de 2011

1. Introdução

Uma pesquisa do CAPI (computador auxiliado por entrevista pessoal) é uma operação complexa que engloba muito mais do que apenas o aplicativo específico para capturar dados. Concedido o CAPI, aplicativo para captura de dados é o coração de todo o sistema, mas como sabemos, uma corrente é tão forte como o mais fraco de seus elos. Assim, devemos preocupar-nos sobre todas as diferentes tarefas que se relacionam directamente ou indirectamente para o objectivo central, a recolha de dados estatísticos de boas, som.

O software desenvolvido para capturar os dados inclui diferentes etapas que abrem o caminho para o objectivo final, principalmente, cuidar-se de aspectos que têm a ver com os seguintes temas:

- Ajudando a determinar que o quadro de design da amostra seja estritamente observado durante o trabalho de campo no sentido de que o número correto de unidades de habitação/domésticos é entrevistado por área cluster ou enumeração.
- Nenhuma unidade de habitação/domésticos ou amostra é duplicado ou ausente.
- Se o exemplo é um exemplo de habitação, ferramentas são oferecidas para entrevistar todas as famílias pertencentes a cada habitação.
- Dentro de cada lar, uma metodologia é seguida para verificar que todas as pessoas elegíveis são entrevistadas (se é utilizada uma amostra de agregado familiar).

Desde que estamos principalmente preocupados com a melhoria da qualidade dos dados recolhidos, todas as actividades que têm um impacto sobre este objectivo devem ser a nossa preocupação. As tarefas que iremos descrever nesta documentação são todos instrumental em atingir este objectivo, embora em algumas ocasiões as etapas descritas não vai mostrar esta prova tão claramente.

As diferentes actividades envolvidas no sistema de controle são classificadas em três grupos diferentes, com base em funcionários que são responsáveis por eles. Todas as actividades realizadas pela pesquisa administrador (SA) também chamado de director de pesquisa, são agrupados juntos e eles devem ser cumpridos antes do segundo grupo, os supervisores, pode começar com suas próprias actividades. Actividades dos supervisores irão pavimentar o caminho para o terceiro grupo, os entrevistadores do inquérito.

O presente sistema foi concebido para obter um exemplo de moradia e devido a alterações finais os planos de amostragem de projecto teve que ser transformado para trabalhar com uma

amostra de agregado familiar, ou simplesmente qualquer tipo de unidade de amostragem. As alterações introduzidas foram apenas superficiais e o mesmo sistema, assim como poderia ser usado no futuro para uma amostra de moradia ou qualquer tipo de pesquisa se baseia em uma amostra.

O sistema de controle é manipulado por um conjunto de menus que incluem as diferentes actividades acima referidas. Esses menus visam ajudar os membros da equipe na execução de suas tarefas e, em muitos aspectos, orientá-los no tempo e da sequência em que essas tarefas devem ser realizadas.

A figura a seguir ilustra a estrutura hierárquica dos três diferentes grupos de funcionários envolvidos em uma pesquisa CAPI.

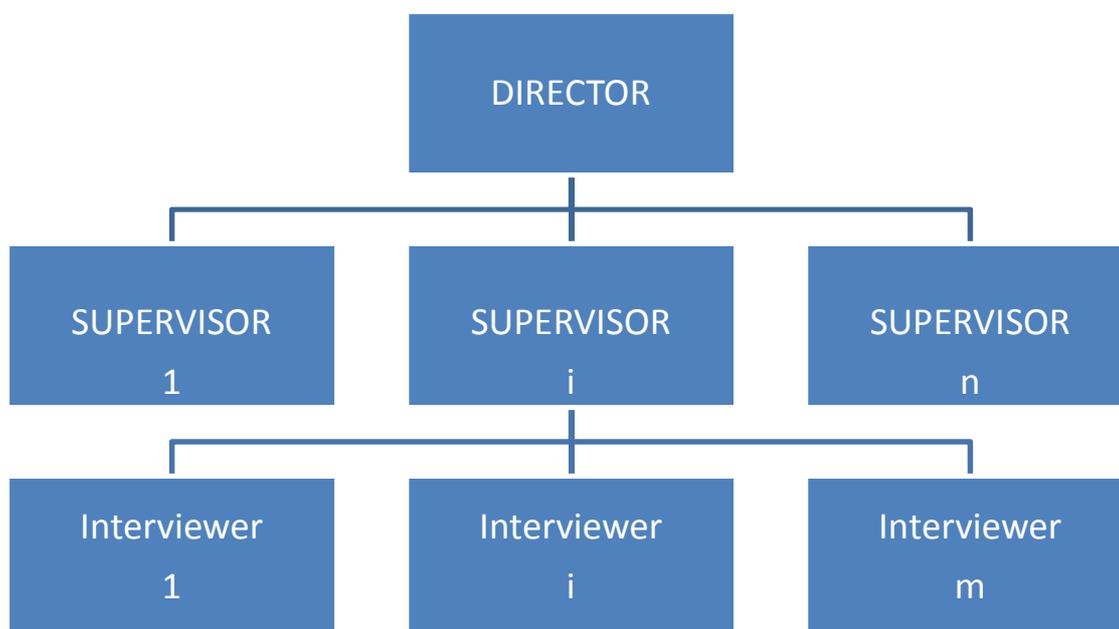


Figura 1: hierárquica estrutura assumido pelo Sistema de controle

O director tem um menu que irá ajudá-los a exercer atribuições do seu director durante toda a duração do projecto. Pode haver dois tipos de Administração: Director de (i) o desenvolvimento, ou também chamado super-director, que é responsável pelo desenvolvimento e manutenção de aplicações de software, a criação dos arquivos de dados de chave usado pelo sistema de controle, a atribuição de trabalho para os supervisores e a manipulação de arquivos de dados de amostra. (ii) o director assistente que tem somente um subconjunto dos direitos do super-director.

Normalmente, as funções de ambos os directores residirá em uma única pessoa, o super-director.

2. Instalação de máquinas

Juntamente com essa documentação, existe algum software básico que tem que ser distribuído para cada máquina que pertencem a três diferentes grupos acima identificados na Figura 1. O

software vem em três arquivos ZIP chamados: "DISTRIBDIRECTOR.ZIP", "DISTRIBSUPERV.ZIP" e "DISTRIBINTERV.ZIP". Como é fácil de imaginar, o primeiro arquivo deve ser copiado para o computador do director, o segundo para máquinas de todos os supervisores e o terceiro para máquinas de todos os entrevistadores.

2.1 Identificação da máquina

O sistema requer que cada máquina tem sua devida identificação e estar associado a um agente. Como temos três tipos de categorias de membros do pessoal, sugerimos que a identificação da máquina (MID) começa com um ou mais letra relacionadas ao grupo que ele pertence. Por exemplo, uma máquina de supervisor pode começar com a letra "S" ou "SUP", mostrando que é uma máquina de supervisor. A parte restante do meados pode ser até a preferência do director; pode ser um código que identifica a máquina no estoque da instituição ou pode ser o código do agente (discutiremos esse conceito no final desta documentação). O fato é que cada máquina pertencentes ao projecto precisa ter uma única MID. Em outros projectos, fizemos rótulos para cada máquina e podemos colocá-lo no computador para facilitar a identificação. Esta será uma identificação física. No entanto, esta identificação deve ser gravado no disco rígido do netbook/notebook/tablet, então ele pode ser acessado pelo sistema de controle.

Vamos tirar partido de uma instalação de Windows usado chamada "Variáveis de ambiente" para definir o meados de cada netbook. Em seguida, mais tarde, o software será capaz de recuperar o MID, sempre que for necessário. Para definir o MID, você pode percorrer as seguintes etapas:

- Clique no botão "Iniciar" do Windows
- Clique em "Painel de controle" do lado direito do Menu Iniciar
- Seleccione a opção "Sistema e segurança"
- Seleccione a opção "Sistema"
- Seleccione o "configurações avançadas do sistema"

Neste ponto, você deve ter a seguinte caixa de diálogo exibida na tela:

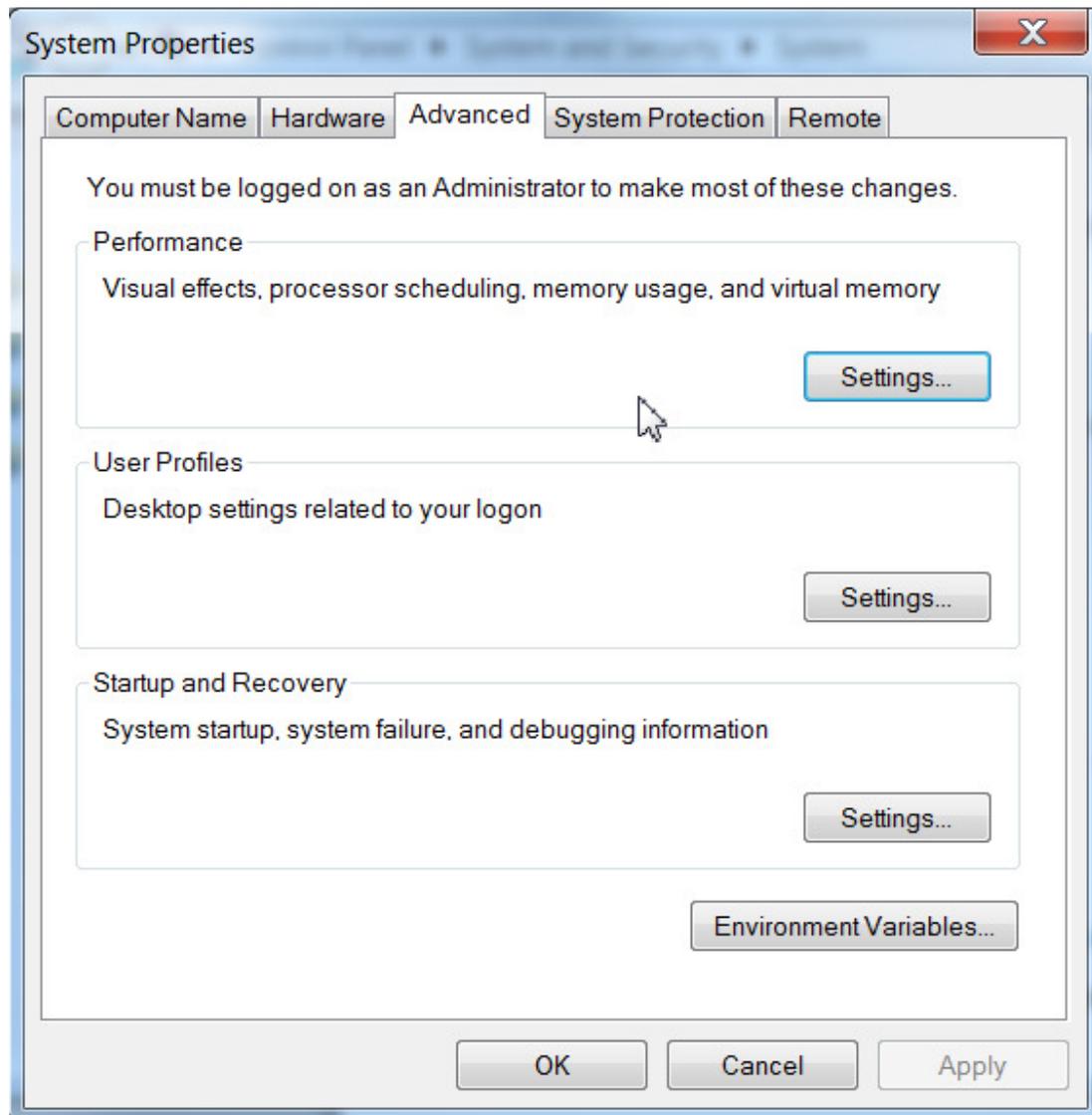


Figura 2: Intermediário passo na definição da variável de ambiente MID

- Clique na aba "Variáveis de ambiente"
- Você deve ter uma caixa de diálogo Nova, como ilustrado abaixo; Observe que os diálogos e os nomes podem variar um pouco dependendo da versão do Windows sua máquina tem

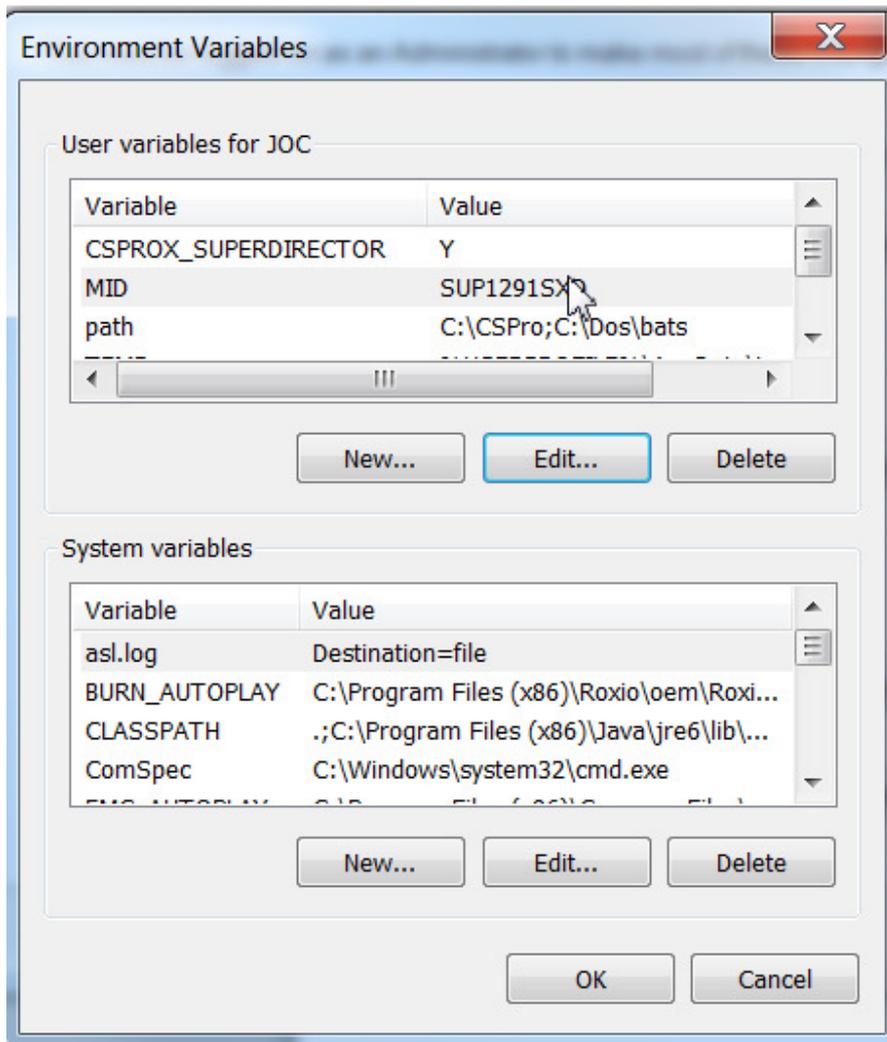
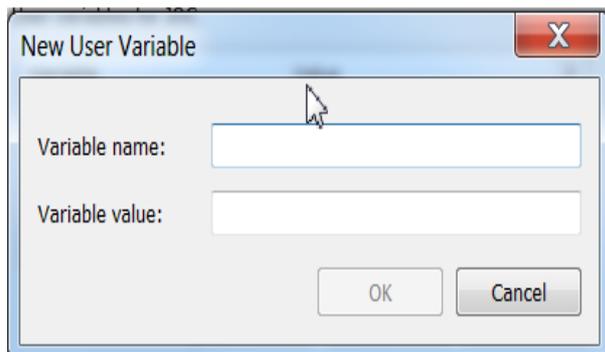


Figura 3: etapa final na definição da variável de ambiente MID

The upper window is the one to define the user variables; as you realize, this machine has already the MID variable defined from a previous project (SUP1291SXD); however, the machine you are working with it's unlikely to have an MID variable defied. Thus, click on the "New" tab on the upper window to define the MID variable. Then, you should get a dialog box like the one below:



Como você pode imaginar, na caixa "Nome da variável" precisamos entrar "MID", que é o nome do nosso id de máquina; na caixa de "Valor da variável", precisamos definir o id de máquina que temos atribuído a esta máquina (ou seja, "SUP1291SXD").

Após concluir essas etapas, a máquina será identificada pelo nome MID que você seleccionou.

2.2 Definição do número IP (Windows 7)

Controle o sistema precisa se comunicar máquinas diferentes a fim de passar informações em toda as etapas da pirâmide indicada na Figura 1 acima. Entrevistadores vão precisar passar dados de suas máquinas para os supervisores e supervisores para o entrevistador, também. Então, se a comunicação local adequada existe, intercâmbio de informações entre o director do projecto e os supervisores devem ser estabelecidas.

Existem várias maneiras de estabelecer comunicação entre dois computadores e o que estamos descrevendo aqui é apenas um deles; no entanto, seleccionámos a opção que se segue porque achamos que é um do mais simples, mais barato e mais seguro, provando ser um método confiável que não deu qualquer problema durante as experiências de muitos países diferentes. O método é a conexão de dois netbooks/computadores usando um cabo de par trançado disponível na maioria das lojas de electrónicos por meio de sua interface física de RJ45, normalmente utilizado para cabos Ethernet. O custo total do cabo é de aproximadamente US\$ 5-10, dependendo do país.

A conexão de rede mais comum será a estabelecida entre cada supervisor e o seu entrevistador. Desde que o tempo necessário para transferir informações entre eles é curto, nós preferimos fazê-lo um entrevistador no momento. Além disso, nem todos os entrevistadores vêm trocar suas informações ao mesmo tempo nem mesmo no mesmo dia.

Para começar, precisamos configurar a placa de rede da máquina dos supervisores para um endereço IP (Internet Protocol) de correcção: "192.168.1.1" e dos entrevistadores rede cartões para "192.168.1.2".

Para fazer isso, siga os passos especificados abaixo:

- Clique no botão "Iniciar" do Windows
- Clique na opção "Painel de controle"
- Seleccione a opção "Rede e Internet"
- Seleccione a opção "Rede e Compartilhamento Center"
- Seleccione "Configurações de adaptador de mudança"
- Clique duas vezes na rede de cabo para acessar as propriedades de rede
- Seleccione o "protocolo de Internet versão 4"
- Definir a IP e máscara de sub-rede, conforme mostrado na figura seguinte

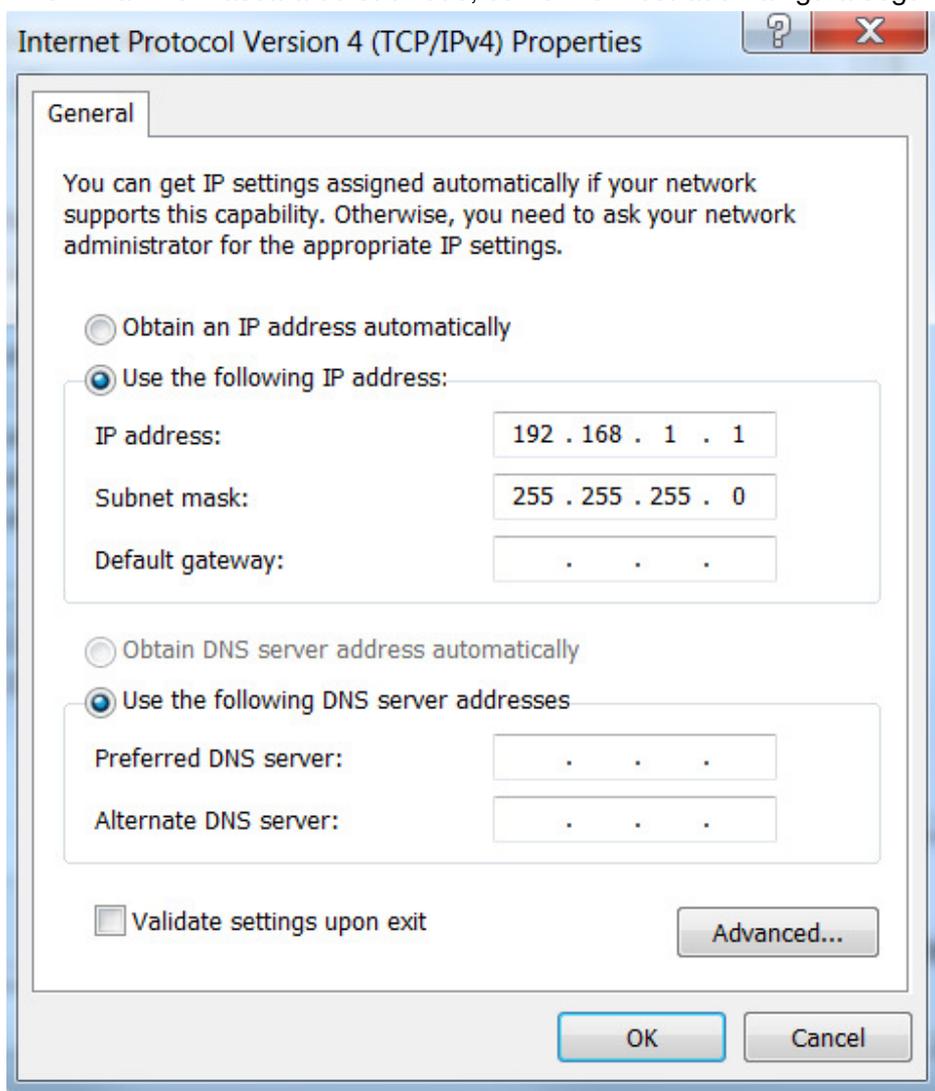


Figura 5: Definição de IP e máscara de sub-rede para a máquina de supervisor

Siga as mesmas etapas para definir o endereço IP e máscara de sub-rede para máquinas dos entrevistadores.

2.3 Estruturas de pasta e arquivos a serem transferidos

Como já mencionado, existem três tipos de máquinas no nosso projecto: o director, os supervisores e os entrevistadores máquinas. Cada um deles precisa ser configurado de forma semelhante, mas com algumas diferenças importantes. Além disso, nós também deve estar cientes que o servidor central, geralmente, um servidor FTP (File Transfer Protocol), desempenha um papel muito importante no sistema de controle e também será analisado.

Dentro as tarefas de configuração de máquina que precisam ser executadas, precisamos estar cientes de que cada tipo de máquina tem um conjunto predefinido de aplicações e arquivos de dados que precisam ser copiados para eles. Também discutiremos sobre os arquivos que precisam ser passadas para cada máquina para garantir a correta comunicação entre eles.

2.3.1 Estrutura de pastas do director

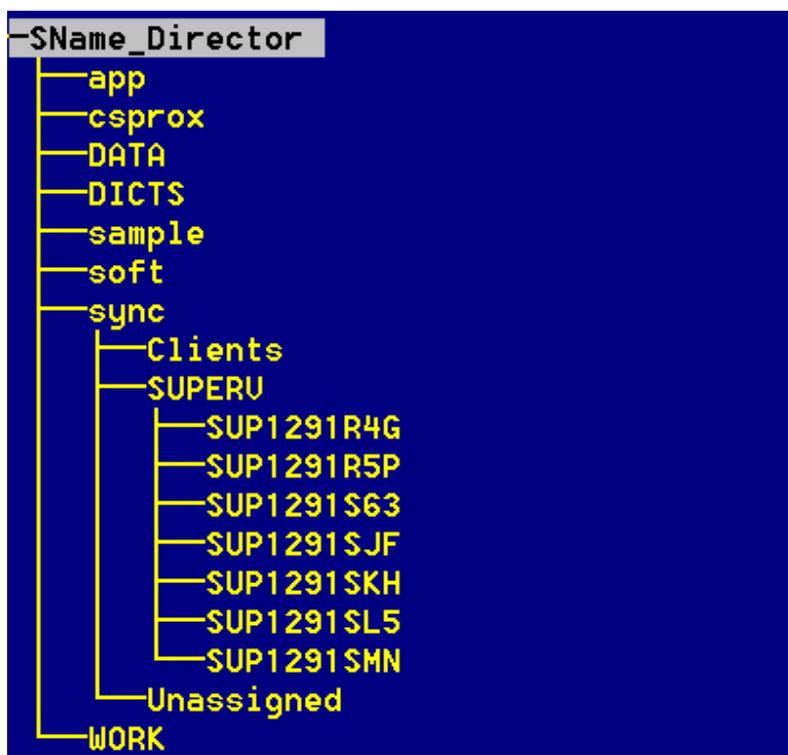


Figura 6: Estrutura de pastas do director

Pasta app: Todos os aplicativos que serão usados pelo director serão inicialmente copiados para esta pasta; incluem arquivos de aplicativo (*.app) formulários ou dados capturar telas (*.fmf), arquivos de mensagens de erro (*.mgf), arquivos de perguntas (*.qsf) e muitos outros. Além disso, alguns arquivos de dados específicos podem ser copiados aqui se alguns aplicativos precisam deles.

CSProX pasta: A pasta CSProX será copiada aqui para certificar-se de que a versão apropriada do sistema é usada. Todos os aplicativos no App acima irão executar directamente o módulo adequado desta pasta.

Pasta de dados: De modo geral, essa pasta irá armazenar os dois arquivos de dados básicos, descritos mais adiante nesta documentação (Teams.dat e Control.dat). Em alguns projectos especiais, algum outro tipo de arquivos de dados também será armazenado aqui.

Dicts pasta: Todos os dicionários de dados usados pelos aplicativos no App serão armazenados aqui; a execução será os dicionários de trabalho que são armazenados juntamente com o pedido em si (App).

Pasta mole: Alguns utilitários de CSProX particulares são armazenados aqui.

Pasta de sincronização: Esta pasta irá armazenar os utilitários de sincronização juntamente com seus respectivos ".Ini "e seus arquivos de chave ser capaz de executá-los (arquivos protegidos). Se a chave não for encontrada, a execução irá falhar. Dois utilitários de sincronização serão usados: (1) o "CSFileSyncClient.exe", que irá sincronizar a máquina do director com o servidor FTP e (2) o "CSFILESyncServer.exe" que, no momento, não tem qualquer função.

A pasta de sincronização tem três subpastas, mas na funcionalidade do director, o único que é activo: a pasta "**Sync\Superv**" que é usada para armazenar as informações de cada supervisor. Os dados armazenados em cada pasta vem da máquina diferente do supervisor e foi armazenados em um arquivo zip e dentro dela, vários dados tenham sido concatenados em um arquivo chamado "AllData.dat". Como é mostrado na Figura 6, sob essa subpasta, uma pasta para cada supervisor é gerada nomeá-los por id de máquina do respectivo supervisor. Hora de que máquina do director é sincronizada com o servidor FTP, dados de cada supervisor do servidor são copiados para a pasta adequada na máquina do director.

Pasta de trabalho: Esta é uma pasta de trabalho usada pelo sistema para executar operações diferentes, como copiar arquivos para diversos fins (zipar, concatenando, etc.).

2.3.2 Estrutura de pastas do supervisor

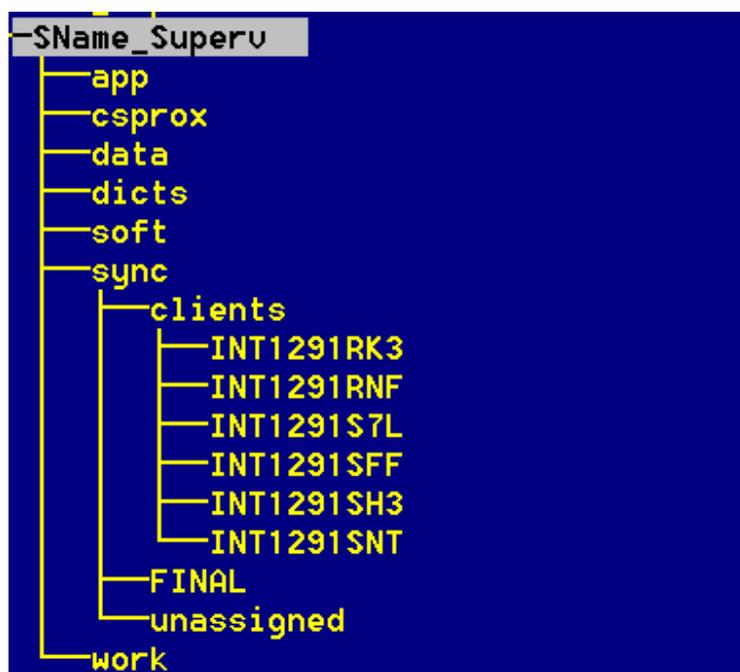


Figure 7: Supervisor's machine folder structure

Estrutura de pastas dos supervisores é muito semelhante do director, e assim, vamos nos referir apenas para as pastas que tem alguma diferença ou particularidade quando comparado a pasta do director.

Pasta app: os aplicativos armazenados aqui foram projectados especificamente para ser usado pelo supervisor e são apenas um subconjunto dos armazenados na máquina do director. A maioria deles nunca é utilizada pelo director, embora eles também são armazenados na pasta o seu App.

Pasta de sincronização: conforme ilustrado na Figura 7 acima, a pasta de sincronização não tem a subpasta "**Superv**" como na máquina do director; em vez disso, ele tem a subpasta "**clientes**" e dentro dela, tem uma subpasta para cada entrevistador da equipe do supervisor. Pastas dos entrevistadores são nomeadas usando que o entrevistador id da máquina e será armazenado todos os dados capturados por cada entrevistador. Pastas dos entrevistadores são criadas pela primeira vez a aplicação de Supervisor Menu é executada ou executada. Observe que os utilitários de comunicação são armazenados na pasta de sincronização, como no caso do director.

Pasta final: Esta é uma pasta usada apenas por supervisores e que irá armazenar todas as entrevistas, modificadas por um supervisor.

Não atribuídos a pasta: Ele armazenará os fólhos ou unidades ids que tem sido atribuídos a um entrevistador; Estas unidades irão adicionar a essas unidades que nunca foram atribuídas de acordo com uma lista de potenciais unidades a ser atribuído e reatribuídos. Essa pasta é usada apenas pelo supervisor.

2.3.3 Estrutura de pastas dos entrevistadores



Figura 8: Estrutura de pastas de máquinas dos entrevistadores

Dados pasta: Além os arquivos "Teams.dat" e "Control.dat" Esta pasta armazena todas as entrevistas capturadas por cada entrevistador.

Pasta de sincronização: Como opor-se a duas pastas já mencionadas, a pasta de sincronização dos entrevistadores não tem quaisquer subpastas; Ele armazena apenas os utilitários de comunicação.

A parte restante das pastas mostradas na Figura 8 tem uma finalidade similar como os respectivos mostrados do director ou pastas de supervisores.

2.4 Os arquivos iniciais (Kernel) copiados para as diferentes máquinas

Cada conjunto de máquinas tem seu próprio kernel que irá fornecer as ferramentas básicas para comunicar as diversas máquinas e, posteriormente, baixar o software completo. O núcleo dos diferentes atores é parte do software que vem com o sistema e deve ser copiado para a máquina de (i) o director/s/s; (ii) as autoridades de supervisão de máquinas e (iii) máquinas dos entrevistadores.

- **Kernel do Director:** o arquivo é um arquivo ZIP "DISTRIBDIRECTOR.ZIP" e ele tem todo o software necessário pelo Director; uma vez que este tem de ser a primeira máquina a ser configurado, o software incluído não é apenas o kernel, mas todos os aplicativos e arquivos de dados. Se ao longo da vida do projecto, as alterações são feitas no software, os arquivos de distribuição (kernel) precisará ser regenerado. Para este efeito, o sistema tem um arquivo de lote chamado "DistribAll.bat" que está localizado na raiz das pastas do director. Este BAT irá gerar os três núcleos automaticamente. No caso particular do kernel do director, situa-se na lista de arquivos que precisam ser compactados em um arquivo de texto chamado "DistribDirector.txt". Se um aplicativo novo tem de ser gerada ou arquivos de dados adicionais são necessários, os nomes de cada arquivo adicional deve ser adicionado à lista em "DistribDirector.txt".

- **Kernel del Supervisor:** o arquivo é um arquivo ZIP chamado "DistribSuperv.ZIP" e ele tem as aplicações mínimas para ser capaz de se comunicar com o servidor FTP. O director vai passar os arquivos de dados e software para o servidor e a partir daí para os supervisores. Assim, o kernel dos supervisores tem o mínimo de software básico para sincronizar com o servidor e baixar todo o software necessário, realmente compactado em um arquivo chamado "ApplicationsforSuperv.zip". A lista de todos os aplicativos que vai ser fechado neste arquivo encontra-se sobre o arquivo de texto "ApplicationsforSuperv.txt" e é usado como guia para gerar o arquivo de zip. Se qualquer arquivo de dados ou aplicativo novo precisa ser adicionado, ele deve ser adicionado à lista de arquivo de texto. Quando a máquina de supervisor sincronizado com o servidor FTP, serão baixados dois arquivos: "ApplicationsforSuperv.zip" e "ApplicationsforInterv.zip". O segundo arquivo será baixado para ser transferido para os entrevistadores ao sincronizar com os supervisores. Observe que a máquina do entrevistador irá nunca comunicar/sincronizar com o servidor FTP e, portanto, ele tem de ser passado por meio de supervisores.
- **Kernel del Entrevistador:** o arquivo é um arquivo ZIP chamado "DistribInterv.zip", tendo o mínimo de software para sincronizar com o supervisor. A lista de aplicação identificando as aplicações de kernel pode ser encontrada no arquivo "DistribInterv.txt". Os arquivos de aplicativos e dados restantes necessários por entrevistadores serão baixados na máquina do supervisor a sincronização entre as duas máquinas. Como mencionado acima, todo o conjunto de aplicativos e dados necessários para os entrevistadores é no arquivo zip "ApplicationsforInterv.zip".

3. Alguns arquivos de dados importantes que precisam ser gerados

O sistema de controle vai contar com arquivo de três dados importantes que terá de ser criado pelo projecto director em dois casos e o sampler ou director demais um terceiro.

3.1 Arquivo de dados equipes

Arquivo de dados das equipes é uma das três fontes muito importantes de informação no quadro da operação do sistema de controle. A finalidade deste arquivo é mostrar a estrutura das equipes; Isso significa que, para cada supervisor, o que são os entrevistadores que farão parte da mesma equipe, mostrando o código, nome, id da máquina, etc. de cada um deles. Além disso, com base no arquivo de exemplo, clusters ou unidades sob investigação serão atribuídas para cada supervisor. A lista de unidades que tem de ser coberto por cada supervisor é inserida e armazenada neste arquivo.

Arquivo de dados das equipes é a primeira definição importante exigida do administrador de projecto (explicado esta documentação). Para ser capaz de fornecer as informações necessárias, é necessário ter um conhecimento profundo da organização de todo o projecto.

Uma equipe será composta de um gerente/supervisor de dados (com experiência no manuseio de computadores, sistema operacional Windows e tendo passado pelo período de treinamento no uso deste software) o número necessário de entrevistadores que passaram por treinamento no uso de netbooks e o aplicativo específico do CAPI. Os clusters atribuídos pelo director do projecto para este supervisor também são especificados aqui.

TEAMS DATA CAPTURE

Supervisor machine Id

Supervisor code

Supervisor's name

Number of Interviewers of this team

INTERVIEWERS

	Interviewer's code	Interviewer's name	Interviewer's machine id
1	<input style="width: 40px;" type="text"/>	<input style="width: 300px;" type="text"/>	<input style="width: 100px;" type="text"/>
2	<input style="width: 40px;" type="text"/>	<input style="width: 300px;" type="text"/>	<input style="width: 100px;" type="text"/>
3	<input style="width: 40px;" type="text"/>	<input style="width: 300px;" type="text"/>	<input style="width: 100px;" type="text"/>
4	<input style="width: 40px;" type="text"/>	<input style="width: 300px;" type="text"/>	<input style="width: 100px;" type="text"/>
5	<input style="width: 40px;" type="text"/>	<input style="width: 300px;" type="text"/>	<input style="width: 100px;" type="text"/>
6	<input style="width: 40px;" type="text"/>	<input style="width: 300px;" type="text"/>	<input style="width: 100px;" type="text"/>
7	<input style="width: 40px;" type="text"/>	<input style="width: 300px;" type="text"/>	<input style="width: 100px;" type="text"/>
8	<input style="width: 40px;" type="text"/>	<input style="width: 300px;" type="text"/>	<input style="width: 100px;" type="text"/>

CLUSTERS ASSIGNED

	From Cluster	To Cluster
1	<input style="width: 60px;" type="text"/>	<input style="width: 60px;" type="text"/>
2	<input style="width: 60px;" type="text"/>	<input style="width: 60px;" type="text"/>
3	<input style="width: 60px;" type="text"/>	<input style="width: 60px;" type="text"/>
4	<input style="width: 60px;" type="text"/>	<input style="width: 60px;" type="text"/>
5	<input style="width: 60px;" type="text"/>	<input style="width: 60px;" type="text"/>
6	<input style="width: 60px;" type="text"/>	<input style="width: 60px;" type="text"/>
7	<input style="width: 60px;" type="text"/>	<input style="width: 60px;" type="text"/>
8	<input style="width: 60px;" type="text"/>	<input style="width: 60px;" type="text"/>

Figura 9: Tela de captura de dados de equipes

O director de projecto precisa ter uma ideia clara da área geográfica onde o supervisor e sua equipe estará cobrindo durante o trabalho de campo para decidir quais clusters devem ser atribuídos. Embora em teoria, os clusters podem ser reatribuídos depois, uma vez que o arquivo de dados de controle é gerado é muito complicado para baralhar aglomerados em torno. Na verdade, cada vez que um cluster é reatribuído, o arquivo de controle terá que ser regenerado.

É importante salientar que o cluster não precisa ser um ponto de amostra; pode ser um fólio ou conjunto de unidades que precisam ser entrevistado, esperançosamente correlacionados a áreas geográficas assim é mais fácil atribuir a um supervisor/equipe.

3.2 Exemplo de arquivo de dados

O arquivo de dados de exemplo pode ser o resultado de um processo algorítmico e estatística para determinar as unidades que terão de ser entrevistado ou simplesmente uma enumeração das unidades que serão visitados e as informações obtidas. São exemplos do tipo de informações que podem ser parte deste arquivo: habitações ou unidades de domésticas, serviços de saúde como clínicas ou hospitais, prefeito, intermediário e diminuir a área geográfica e administrativo, local de tipo urbano/rural, endereço, etc.

Este arquivo é originado do amostrador ou o director de pesquisa e ele será usado para gerar o segundo arquivo de dados que será nosso companheiro para a operação de todo trabalho de campo, o arquivo de controle.

3.3 O arquivo de controle

Este arquivo é o coração deste software e será usado e actualizado constantemente através da cadeia de diferente eventos.

A geração deste arquivo importante é realizada usando as informações de design de amostra juntamente com o arquivo de dados de equipes (já descrita no ponto 2.3) como entradas, produzir o arquivo de controle de dados como saída. As informações armazenadas neste arquivo são relevantes para as finalidades principais do sistema e, portanto, ele vale a pena gastar algum tempo descrevendo, pelo menos, o mais importante deles.

Haverá um registo contendo informações mostradas na Figura 7, para cada unidade da amostra ou arquivo lista descrevendo as unidades que serão entrevistadas.

O cluster e o identificador devem ser interpretado como uma combinação única de identificar uma unidade. Como foi mencionado antes, pode ser um fólio que irá produzir uma identificação exclusiva.

Os níveis geográficos link a unidade com área geográfica específica. Uma vez que esta informação é o mais geral possível, o nível 1 pode ser, em algum lugar, o nível de "Estado" e nível 2 pode ser a "província". O número de níveis geográficos será algo que é definido pelo projecto/país e se apenas duas áreas são necessárias, os outros dois não devem ser aplicáveis.

Main Info of the Control Data File

Cluster/Folio		Identifier	
<input type="text"/>		<input type="text"/>	
Geographical level 1	<input type="text"/>	Supervisor in charged of this unit	<input type="text"/>
Geographical level 2	<input type="text"/>	Interviewer assigned to this unit	<input type="text"/>
Geographical level 3	<input type="text"/>	Number of visits	<input type="text"/>
Geographical level 4	<input type="text"/>	Interview status of last visit	<input type="text"/>
Urban/Rural	<input type="text"/>		
Date when unit was assigned to interviewer	<input type="text"/>	Date when received by interviewer	<input type="text"/>
	<input type="text"/>		<input type="text"/>
		Date of the interview	<input type="text"/>
			<input type="text"/>
		Date returned to supervisor	<input type="text"/>
			<input type="text"/>

Figura 10: Informação de principal do arquivo de dados de controle

Novamente uma vez que as especificações variáveis tentam ser o mais geral possível, definimos o código "Supervisor" e o "código do entrevistador" como uma variável de cinco dígitos. No entanto a maioria das vezes o código será um dois ou três algarismos que sugerimos abaixo:

Código de Supervisor	Código do entrevistador
100	101-107
200	201-206
....
900	901-909

O "Status da entrevista da última visita" é muito importante, pois ele será usado para mostrar um entrevistador e/ou o seu supervisor o status geral do seu trabalho para um cluster específico.

- Data quando a unidade é atribuída pelo supervisor para o entrevistador: é a primeira acção do supervisor com o seu entrevistador e mostrará o momento para começar a seguir até o unidade-entrevistador.
- Data em que a atribuição é recebida pelo entrevistador: os dados acima foram apenas a acção/decisão da autoridade para atribuir a unidade para o entrevistador; Esta data é o momento quando o entrevistador começa a atribuição na sua máquina. De agora em diante, a unidade específica vai aparecer na lista a ser seleccionado para ser entrevistado.
- Data da entrevista: Esta é a data em que a entrevista é executada pelo entrevistador.

- Data retornada ao supervisor: Este é o momento quando os dados da entrevista são passados para o supervisor.

Com quatro datas acima, mais a identificação da unidade, o supervisor e identificação do entrevistador, estamos na capacidade de seguir para cima e dizer os director de pesquisa/supervisores o status de qualquer unidade a qualquer momento.

4. Sistema de controlo a orientado por menus

Como mencionado anteriormente, o sistema de controle inteiramente é tratado por menus orientados para três tipos de funcionários, o director de pesquisa, os supervisores e entrevistadores. Cada um deles tem um número bem definido de actividades que são retratados em seus respectivos menus.

Como analisamos os menus diferentes, vamos entender melhor os conceitos e finalidades do sistema de controle. Vamos começar analisando o menu do director, explicando a opção por opção as diferentes tarefas que ele pode e deve fazer.

4.1 Super-diretor do menu

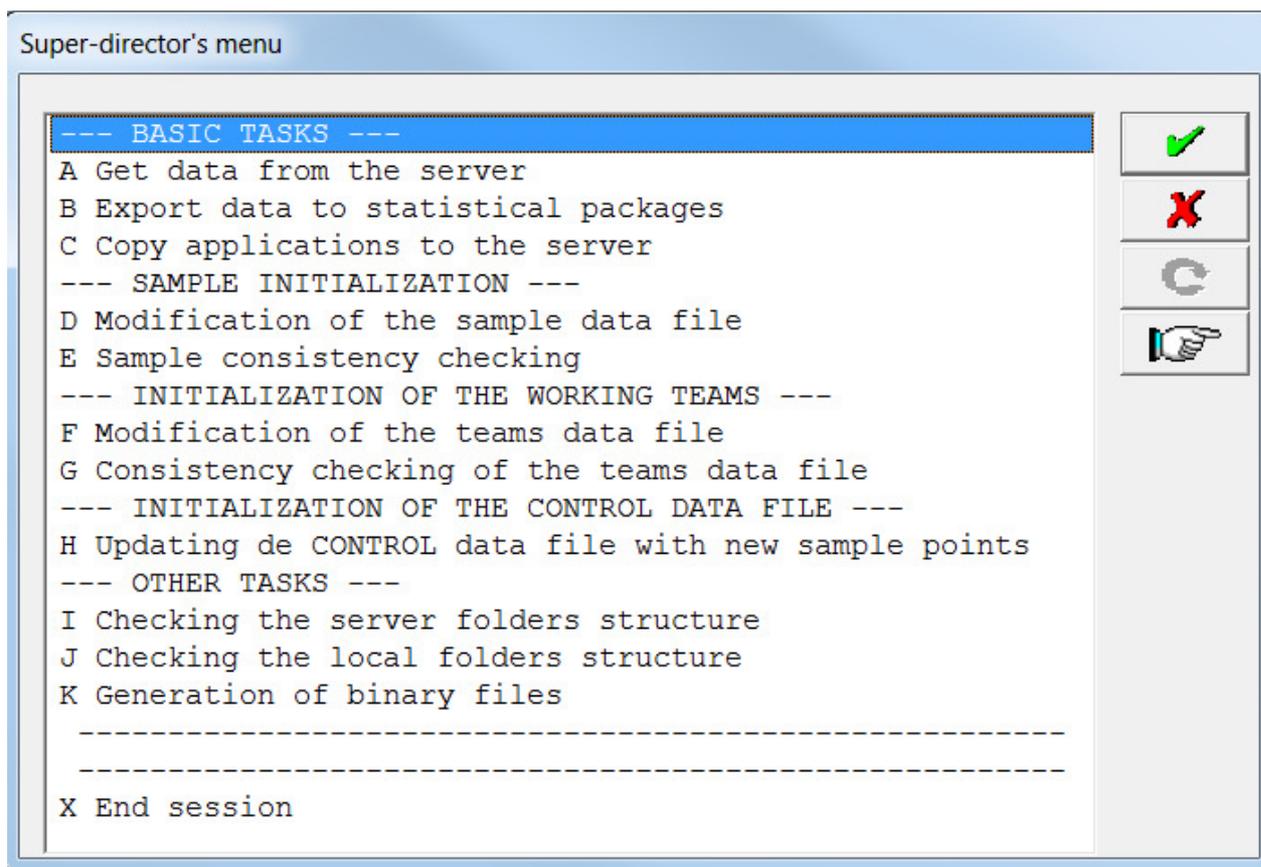


Figura 8: Menu de Super-diretor

Para ser capaz de executar este menu, a máquina deve ter a variável de ambiente "CSPROX_SUPERDIRECTOR = Y". Assim, ao mesmo tempo que o director define a variável MID, ele ou ela também deve definir esta variável. Máquina do nenhum outro projecto deve ter essa variável de ambiente definido. Se a variável não está definida, o subconjunto do menu para o director regular será usadas (opções A, B e X).

Opção A: obter dados do servidor

Normalmente, os dados colectados por entrevistadores são passados às máquinas dos supervisores e de lá para o servidor central. Dizemos normalmente porque em muitos países a internet a comunicação não é tão bom em áreas remotas, como é nas grandes cidades. Assim, nessas situações, é preciso criar uma alternativa para passar as informações de máquinas dos supervisores para o servidor central.

O servidor central, portanto, é a máquina que armazena juntos todos os dados recolhidos por entrevistadores e usando esta opção, o director é capaz de passar todos os dados para o seu computador.

Quando o director selecciona esta opção, arquivos de dados de cada supervisor serão baixados para máquina do director e o concatenados em um arquivo chamado "AllData.dat".

Opção B: exportar dados para pacotes estatísticos

O arquivo de dados "AllData.dat" é exportado para o SPSS, SAS e Stata.

Opção C: copiar aplicativos e arquivos de dados para o servidor

Como mencionado antes, o director é responsável pelo desenvolvimento de aplicativos para ser usado no projecto CAPI e também é responsável pela geração das equipes.DAT e o CONTROL.DAT. Uma parte importante do trabalho será para passá-los para as máquinas apropriadas na hierarquia mais tarde.

A maneira de distribuir os aplicativos e dados para máquinas os supervisores requer que eles sejam transferidos para o servidor central pela primeira vez. A partir daí, a informação precisa ser passado de supervisores (menu de supervisores Veja mais tarde) e para os supervisores para seus respectivos entrevistadores (menu dos entrevistadores de ver).

Esta opção irá copiar todos os aplicativos e arquivos de dados necessários para o servidor para que eles podem ser distribuídos para máquinas dos supervisores. Existem dois arquivos que incluem o conjunto completo de aplicações necessárias: "ApplicationsforSuperv.zip" e "ApplicationsforInterv.zip". Como o nome indica, a primeira inclui todos os aplicativos necessários por máquinas dos supervisores e a segunda por entrevistadores.

Note que o servidor não distribuir os arquivos para os entrevistadores. Isto só pode ser conseguido por meio de supervisores.

Opção D: modificação do arquivo de dados de exemplo

Normalmente, o sampler irá fornecer os dados de exemplo em uma planilha ou em qualquer outro formato que deve ser adaptado para um arquivo de dados CSPro. Em qualquer caso, esta é uma operação que é dependente da instalação e terá de ser preparado e concluído pelo director.

Opção E: verificação da consistência de amostragem

Novamente, desde que a amostra é dependente da instalação, essa opção será necessário ser concluída e refinado pelo director.

Opção F: modificação do arquivo de dados de equipes

Uma explicação completa sobre o "**Teams.dat**" foi fornecida antes na secção 3.1. Esta opção, em seguida, irá permitir que o director de projecto para criar e modificar esse arquivo de dados importantes.

Opção G: consistência verificação do arquivo de dados de equipes

Esta é uma opção muito importante que irá verificar que nenhum cluster é duplicada, que nenhum id de máquina é duplicado e, em geral, que as equipes é de arquivo de som e pronto para ser usado.

Opção H: Actualizando o arquivo de dados de controle

Se novos clusters são adicionados à amostra após o arquivo de controle já gerou, essa opção permitirá a incorporá-las no arquivo de dados de controle?

Opção I: Verificando a estrutura de pastas do servidor

Esta opção irá verificar que todas as pastas do servidor FTP estão definidas correctamente como especificado no "DirectorsParameters.txt" ou o "DirectorsParametersInternet.txt". Arquivo de texto que será usado vai depender se o servidor FTP é remoto (acessado através da internet) ou não. Além disso, esta opção irá criar todas as subpastas sob "Sync\Superv", um para cada supervisor, conforme definido no arquivo de dados de equipes.

Opção J: Verificando a estrutura de pasta local

Isso é muito parecido com a opção acima, mas é realizado em máquina do director.

Opção K: Geração de arquivos binários

Esta opção irá produzir todo o conjunto de arquivos binários para o aplicativo de CAPI. Do ponto de vista de segurança, é importante distribuir apenas os arquivos binários para que o aplicativo não pode ser mudado ou alterado por entrevistadores nem os supervisores. Esses arquivos devem ser regenerados toda vez que qualquer um dos arquivos de dados do aplicativo é modificado.

4.2 Menu a autoridades de supervisão

A Figura 9 ilustra as diferentes opções que o supervisor tem de realizar as tarefas importantes necessárias bloqueá-lo.

Opção A: Executar o servidor

Como foi mencionado antes, há dois utilitários de comunicação distribuídos com o sistema de controle, um é o "CSFileSyncServer.exe" e o segundo, "CSFileSyncClient.exe". Ambos devem ser localizados e armazenados na pasta de sincronização de cada máquina deste projecto. O servidor deve ser executado somente quando o supervisor quer iniciar a sincronização com os entrevistadores; para se conectar com o servidor central, G deve ser usada.

O arquivo "CSFileSyncServer.ini" tem importantes parâmetros definidos, especificando a porta que será usada na comunicação e os usuários que serão capazes de se conectar. Este arquivo é projecto e instalação dependente e assim, ele precisa ser devidamente definido de acordo com as necessidades locais.

Os usuários que precisam ser definidas aqui são o supervisor e todos os entrevistadores, trabalhando para a equipe do supervisor.

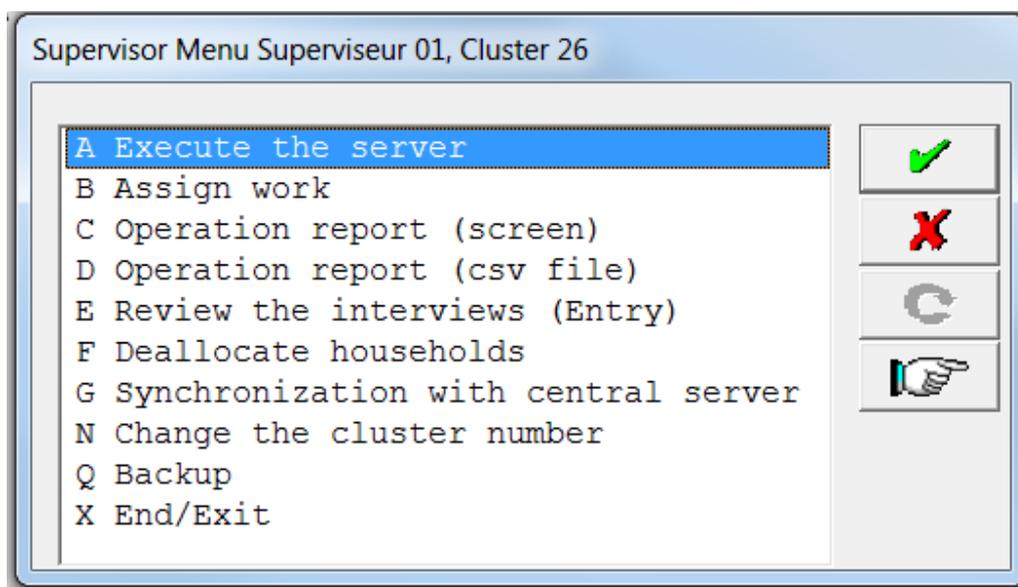


Figura 9: Menu do Supervisor

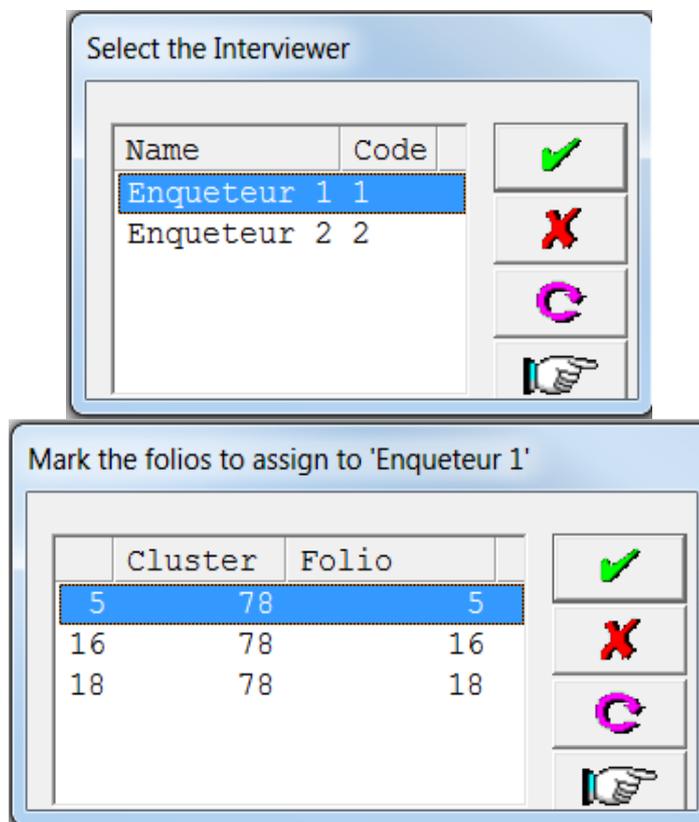
Opção B: Trabalho de atribuir

Figura 10: caixas de diálogo a atribuição de trabalho para entrevistadores

O arquivo de dados de equipes tem, para cada equipe/supervisor, uma lista de todos os clusters, o director atribuído a ela. No entanto, esses clusters meu incluem uma ou mais folhas que fazem parte do trabalho a equipe tem a entrevista mas precisam de ser atribuído a um ou mais entrevistadores. Como o progresso do trabalho de campo, a lista de fólhos e clusters pendentes ser entrevistado fica menor; o trabalho de campo para que o supervisor terá fim somente quando a lista é nula.

O supervisor irá atribuir trabalho aos entrevistadores periodicamente com base na velocidade que eles executam suas funções. Para atribuir o trabalho a eles, o supervisor deverá executar duas etapas: para seleccionar todos os entrevistadores pertencentes à equipe o um específico para atribuir trabalho (primeira caixa de diálogo da Figura 10); Depois que o entrevistador tem sido adequadamente seleccionados, então o sistema irá mostrar a lista de fólhos para o cluster seleccionado que não receberam ainda. No caso particular do exemplo ilustrado na Figura 10, o cluster seleccionado é 78. Observe que o primeiro passo é executar o "Superv_Menu" será a selecção do cluster que deseja trabalhar com (consulte opção N, "Alterar o número de cluster").

Opção C: Relatório de operação

Cada supervisor precisa saber o comportamento dos entrevistadores em relação a atribuição de trabalho. O sistema é capaz de fornecer, com base no status das entrevistas gravadas no

arquivo de dados de controle, a quantidade de trabalho realizado ou pendente pelo entrevistador ou por toda a sua equipe, pelo cluster ou para todos os clusters. Para solicitar relatórios específicos, o sistema fornece as seguintes alternativas de que o supervisor pode fazer a selecção:

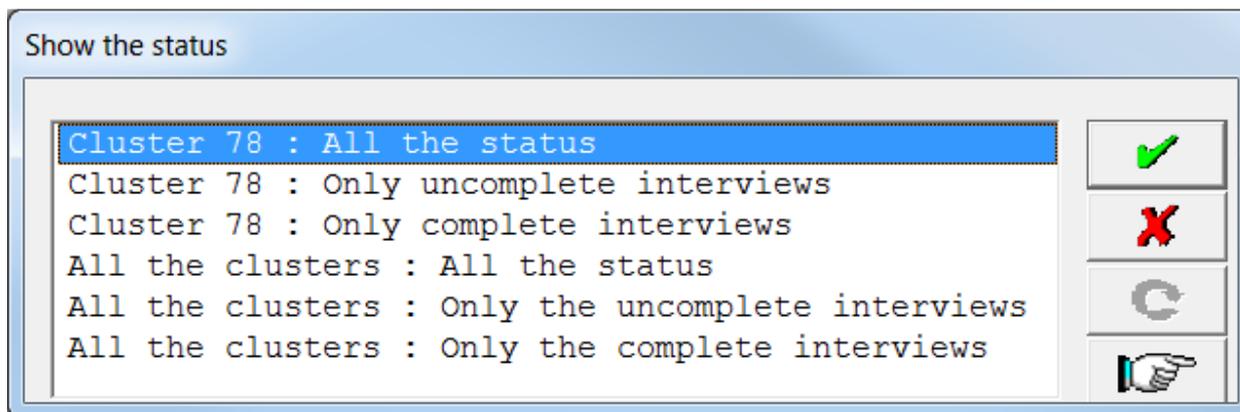


Figura 11: Relatórios de status de trabalho diferentes

Como no exemplo anterior, estamos trabalhando com cluster actual é 78. Se precisamos obter o relatório em um cluster diferente, há a opção para alterar o cluster actual. As três primeiras opções acima se referem a (i) recebendo o status de todos os fólhos/entrevistas para cluster de 78; (ii) ficando apenas entrevistas incompletas e (iii) ficando apenas as entrevistas completas. Os mesmos relatórios podem ser obtidos para todos os clusters em vez de cluster apenas 78.

Opção D: Relatório de operação (arquivo csv)

Esta opção realizar exactamente o mesmo que a opção "C" acima, mas em vez de ser dirigida para a tela, é dirigido a um "Comma-Separated Value", também conhecido como um arquivo de dados csv. Este arquivo de saída pode ser facilmente importado para o Excel, para fins de análise.

Opção E: Rever as entrevistas (entrada)

Esta opção permitirá que os supervisores verificar se as entrevistas com os mesmos dados capturar o entrevistador usado no campo de aplicação.

Opção F: Rever as entrevistas em lote

Aqui, o supervisor pode verificar uma ou mais entrevistas, usando o mesmo aplicativo para capturar os dados de entrevista mas em lote de modo; erros detectados são impressas em um relatório, que é conveniente quando se trata de analisar a qualidade do trabalho realizado pelos entrevistadores.

Normalmente, o supervisor vai querer verificar mais de uma entrevista na época; para definir o universo que será verificado, o sistema fornece uma caixa de selecção semelhante ao relatório oferecido para inspeccionar o trabalho dos entrevistadores. Assim, o supervisor pode dimensionar o tamanho do arquivo em lote para verificar de acordo com as necessidades.

4.3 O menu de entrevistadores

O último menu de sistema fornece todas as tarefas que esperamos que os entrevistadores para realizar de forma simples e organizada. A figura a seguir ilustra as funções que o entrevistador deve para executar:

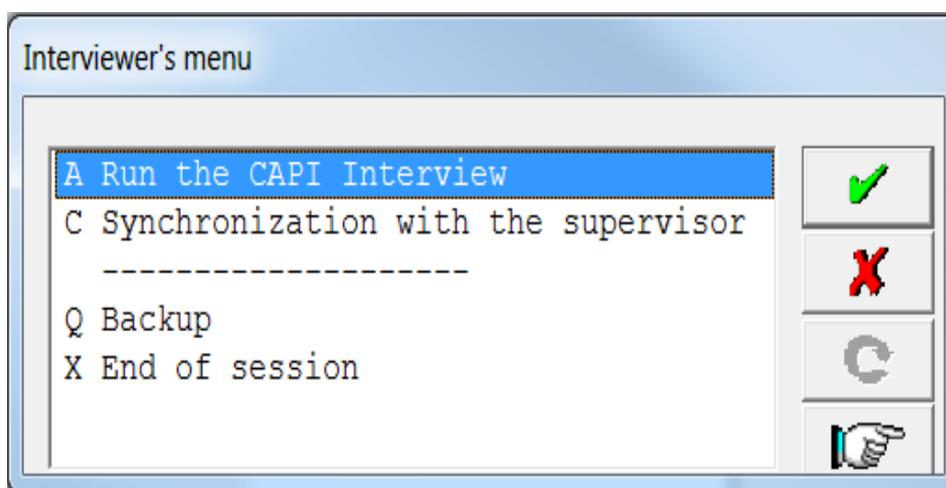


Figura 12: Menu do entrevistador

Opção A: Executar a entrevista CAPI

Esta opção permite que o entrevistador iniciar o aplicativo de CAPI, ou uma nova entrevista com a continuação da entrevista que tinha sido interrompida anteriormente. Desde que um entrevistador pode ter muitas unidades/famílias atribuídas pelo supervisor, ele deverá seleccionar qual unidade do que as atribuídas será entrevistada em seguida. A seguinte caixa de diálogo será aberta para mostrar as unidades atribuídas (cluster e número de folio/unidade/domésticos), as identificações geográficas e o nome do contacto/pessoa na unidade para ser entrevistado e o status da entrevista.

Como visto na figura abaixo, este entrevistador tem quatro unidades, todas elas com um status de entrevista ou resultar zero (o que significa que a entrevista ainda não foi iniciada). Então, o entrevistador tem a alternativa para seleccionar a unidade que é mais conveniente e iniciar a entrevista. Seleccionando o número desejado, o sistema irá chamar o aplicativo de CAPI e iniciará a entrevista para que a unidade específica.

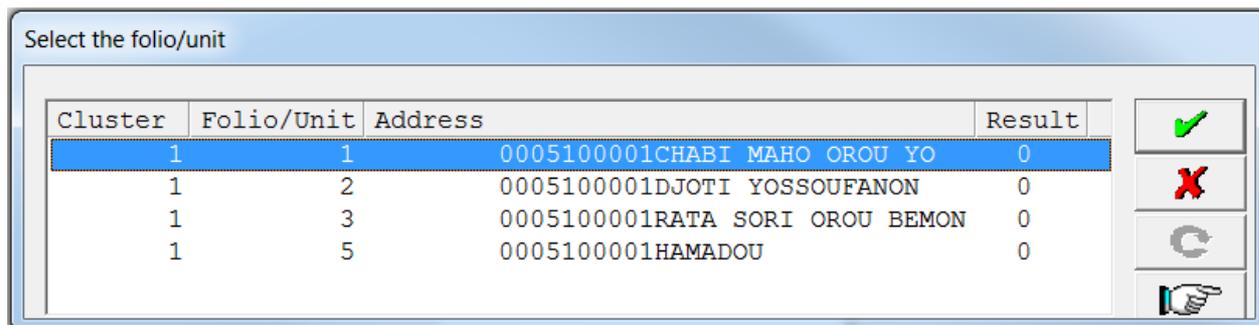


Figura 13: Caixa de diálogo para seleccionar a unidade a ser entrevistado

Lista de clusters e números são extraídos do arquivo de dados de controle e alguns elementos armazenados nesse arquivo são copiados para a primeira tela do aplicativo CAPI (ou seja, supervisor e entrevistador códigos, códigos geográficos, etc.).

Opção B: Sincronização com o Supervisor

Uma vez por dia ou em dias alternados, o entrevistador deve entrar em contacto com o supervisor para realizar as seguintes tarefas:

- Transferência de dados a entrevista se reuniram durante a viagem para o supervisor com duas finalidades: (i) a seguir o até mais tarde para o pessoal responsável para analisar a qualidade dos dados possa detectar qualquer problema relacionado com a aplicação de CAPI ou aos entrevistadores que podem distorcer a realidade sendo investigada; (ii) tornando possível a fusão de arquivos de diferentes dados parciais recolhidos por cada entrevistador, facilitando a criação de um arquivo de dados mestre.
- Receber do supervisor a atribuição de trabalho para a próxima viagem.
- Atualização do software do CAPI no caso em que as alterações foram feitas.

Para sincronizar com o supervisor fazer certeza que (1) o cabo para comunicar-se ambas as máquinas é viciado em ambas as extremidades; (2) Certifique-se que não há nenhum outro entrevistador conectado à máquina supervisor; (3) Certifique-se de que a máquina do supervisor está executando o servidor FTP e portanto, está pronta para receber seus dados.

Opção Q: Backup

É aconselhável você, uma vez por dia, executar um backup completo de seus dados. Desta forma, na eventualidade de qualquer coisa indesejável acontece a máquina do entrevistador, vai haver nenhuma perda ou, pelo menos, um mínimo de perda de dados se arrepender.

O dispositivo de backup, recomendamos usar vai ser um disco flash ou, também chamado, pen drive. Normalmente, a letra de dispositivo atribuída a este dispositivo removível depende da máquina e o número de periféricos instalados; no entanto, comumente máquinas dos

entrevistadores não tem quaisquer dispositivos montados pois são equipados com o mínimo para garantir que a operação de recolha de dados CAPI podem ser alcançados com êxito. Assim, podemos estar certos de que, na maioria dos casos, a letra atribuída à unidade de caneta será "E". No caso de era uma letra diferente, o sistema vai dar o entrevistador a chance de alterá-lo.

Opção X: fim de sessão

Seleccionando esta opção, o sistema irá encerrar a sessão do entrevistador. Essa opção normalmente deve preceder a final do supervisor da sessão, então tudo pode terminar normalmente.