



STATISTICS  
DENMARK



Statistisk sentralbyrå  
Statistics Norway



Statistiska centralbyrån  
Statistics Sweden

MZ:2005:11

## Relatório da missão de curta duração sobre formação em CPro

De 10 a 26 de Junho de 2005

### TA for the Scandinavian Support Program to Strengthen the Institutional Capacity of the National Statistics, Mozambique

*Julio Ortúzar*

**SERPRO S.A.**

Benjamín 2935 Of. 302  
Santiago, Chile 755-0048



Instituto Nacional de Estatística

## RESUMO EXECUTIVO

O principal objectivo desta missão era a capacitação institucional na área de utilização do software CSPro, com particular incidência no controlo de consistência e batch. Durante as duas semanas do workshop, o qual eu chamaria de treinamento no local de trabalho, foram revistas as principais características de CSPro, inclusive as interfaces gráficas, a exportação de dados de CSPro para SPSS/STATA/SAS assim como a produção rápida de tabelas cruzadas. Com estas tabelas pode-se controlar, atempadamente, erros provenientes do trabalho de campo e permitir dessa forma a detecção atempada de problemas que de outra forma poderiam ter um impacto negativo nos dados adquiridos.

Ênfase foi dada à planificação e controlo da operação do inquérito. Apesar de controlos de consistência serem muito importantes para a produção de dados consistentes os quais facilitarão mais tarde a análise dos mesmos, há outro tipo de erros com um impacto maior de precisão em termos de representatividade dos dados e que requerem portanto técnicas diferentes para a sua detecção e correcção. Assim, após a identificação de tais erros, tempo foi disponibilizado para o desenho de estratégias para descobri-los e corrigi-los. Os tipos de erros a que me refiro são principalmente (i) posicionamento incorrecto de UPA's (Unidades Primárias de Amostragem) através da obtenção/entrada de identificação errada (por exemplo: a província/distrito errados), (ii) UPAS em falta e (iii) duplicação de UPA's. Estes três tipos de erros tendem a distorcer a representatividade dos dados em diferentes formas mas o resultado é o mesmo: uma vez tabulados ou convertidos os dados micro em dados macro, estes são difíceis de perceber pese embora os mesmos distorcerem ou apresentem resultados tendenciosos na medida em que dependerão da frequência em que tais erros acontecerem e como estes irão compensar-se mutuamente.

O progresso registado pelo pessoal de processamento de dados do INE foi evidente, tomando em consideração o curto período da missão. O progresso deve ser medido em diferentes áreas tais como: (i) tomar conhecimento das novas características de CSPro, as quais foram revistas e utilizadas imediatamente no Inquérito à Força de Trabalho (IFTRAB) o qual está sendo actualmente processado; (ii) análise sobre como as características de CSPro em geral (antigas e novas características) podem ser utilizadas para diferentes situações do processamento de inquéritos para atingir resultados específicos (por exemplo: ligação de diferentes questionários para o processamento de controlos de consistência entre eles); (iii) sensibilização do pessoal sobre os diferentes tipos de erros que os mesmos devem resolver para melhorar a qualidade de dados do inquérito.

O progresso foi medido e observado por alguns técnicos do INE, demógrafos assim como diferentes autoridades, inclusive um perito do Programa Escandinavo. O teste de avaliação foi efectuado em forma de demonstração da aplicação de entrada de dados, desenvolvida por técnicos informáticos do INE, a qual incluiu uma vasta gama de controlos de consistência, tanto num questionário como entre questionários diferentes e mais importante ainda o controlo de validação de identificação de UPA's com base no ficheiro do desenho amostral. Seguidamente, os mesmos controlos de consistência foram corridos no modo "batch" para testar os dados já digitados e desta forma obter-se uma ideia sobre a qualidade dos dados. De acordo com a audiência, os

resultados obtidos foram impressionantes, considerando que a aplicação não possuía senão testes de intervalos.

Não obstante o evidente progresso feito, há ainda um longo caminho a percorrer em termos de capacitação institucional. Seria pretensioso considerar que em apenas duas semanas de seminário seria possível cobrir a maior parte de teoria sobre processamento estatístico de dados. Contudo, pelo menos, os formandos tiveram a oportunidade de experimentar e adquirir experiência sobre diferentes tipos de erros que podem distorcer a representatividade do universo de dados em estudo. As minhas recomendações sobre passos a seguir são os seguintes:

1. É de grande importância criarem-se equipas de trabalho que combinem pessoal de diferentes capacidades e conhecimento (por exemplo: demógrafos, sociólogos, estatísticos e informáticos) para o desenho de estratégias e planos para estudos futuros. A integração desta equipa tem por objectivo “atacar” os vários problemas que podem ocorrer nas diferentes fases do inquérito/censo começando com o desenho do questionário, desenho da amostra – caso seja aplicável –, formação de inquiridores, trabalho de campo, etc. O resultado final do trabalho desta equipa deve ser um documento escrito, contendo responsabilidades claramente definidas e no qual todas as estratégias para tornar possível a detecção atempada de numerosos problemas que certamente irão acontecer.
2. Implementação de um sistema de controlo. Um inquérito e mais importante ainda, um censo devem ser escrutinados e monitorados por um sistema de controlo, no qual cada pequena unidade – UPA ou área de enumeração (AE) – é acompanhada ao longo das diferentes fases da operação do inquérito/censo. O ficheiro de controlo, onde toda esta informação é gravada, deveria acomodar, pelo menos, a seguinte informação: (i) um registo para cada unidade (AE ou UPA) onde toda a informação relacionada será gravada; tal informação pode ser de duas classes, nomeadamente informação preexistente proveniente tanto do desenho amostral, cartografia e/ou pré/censo, para o caso de censos e informação actual compilada pelos dados obtidos. Isto permite que estatísticos e outros especialistas analisem a validade dos dados obtidos em comparação com os previstos. (ii) um cabaz possa ser gravado para cada fase da operação do inquérito e AE/UPA, data após finalização, etc. Esta simples informação permite um estrito acompanhamento do ponto de situação da operação e ao mesmo tempo responde à maior parte dos erros relacionados com UPA's e/ou AE's, anteriormente mencionados. Relatórios baseados no ficheiro de controlo têm também por objectivo fornecer, à gestão de topo do INE, uma imagem clara do ponto de situação ou grau de realização das diferentes fases da operação do inquérito/censo.
3. Um importante “output” do documento mencionado no ponto 1 deve consistir num ficheiro detalhado de especificações do ficheiro de controlo descrito no ponto 2 e um conjunto completo de regras de controlo de consistência e validação a serem aplicadas sobre o aplicativo de entrada ou captação de e no aplicativo de correcção de dados. Uma vez concluídas estas actividades, o INE pode confiar na qualidade dos dados que o mesmo irá futuramente analisar.

## 1. INTRODUÇÃO

Este relatório foi elaborado por Julio Ortúzar, director executivo da SERPRO S.A. e refere-se à missão levada a cabo entre os dias 10 e 26 de Junho de 2005. O principal objectivo desta missão foi o de ministrar um curso avançado de CSPro, tomando como base para todos os exemplos práticos dois questionários diferentes, nomeadamente o Inquérito à Força de Trabalho (IFTRAB) actualmente em curso e com cerca de 50% dos dados obtidos já digitados e um inquérito agrícola, cujo processo de levantamento de dados está previsto para Agosto de 2005.

Os participantes no curso tinham frequentado pelo menos um curso anterior ao presente e portanto estes possuíam conhecimento e experiência no uso do software. Tomando como base este pressuposto, o curso incluiu a revisão das novas características do sistema, com ênfase na aplicação imediata dos novos conceitos e características nos respectivos inquéritos.

De acordo com os termos de referência, dois problemas principais haviam sido identificados pelo pessoal do Departamento de Informática e Sistemas de Informação (DISI) do INE: treinamento em técnicas avançadas de CSPro com o propósito de acelerar o controlo de consistência e o controlo de lotes.

O IFTRAB, actualmente na fase de levantamento de dados, possui três questionários: (i) informação do agregado familiar, incluindo informação sobre características demográficas dos membros assim como as necessárias ligações descrevendo a composição ou estrutura do agregado familiar; (ii) o questionário principal para pessoas de 6 anos ou mais de idade, no qual a maior parte das perguntas sobre emprego estão inseridas; e (iii) o questionário infantil para pessoas entre os 7 e os 17 anos de idade que tenham trabalhado nos últimos 7 dias.

A maior parte dos problemas identificados pelo pessoal do DISI foram originados pelos três questionários uma vez que os mesmos não sabiam estabelecer a ligação entre uma pessoa no questionário 2 com a informação correspondente à mesma pessoa no questionário 1. Esta situação tornava difícil realizar controlos de consistência entre informação correspondente a diferentes questionários e de igual modo filtros ou saltos, usando informação de diferentes módulos, não eram executados correctamente.

Este relatório contém os pontos de vista do(s) consultor(es), os quais não correspondem necessariamente aos pontos de vista da DANIDA ou do INE.

## 2. ACTIVIDADES REALIZADAS DURANTE A MISSÃO

Contrariamente à maior parte das missões, esta era muito específica e tinha por objectivo transmitir conhecimentos sólidos de CSPro ao pessoal do DISI, capacitando-o desta forma a processar, no futuro, qualquer tipo de inquérito.

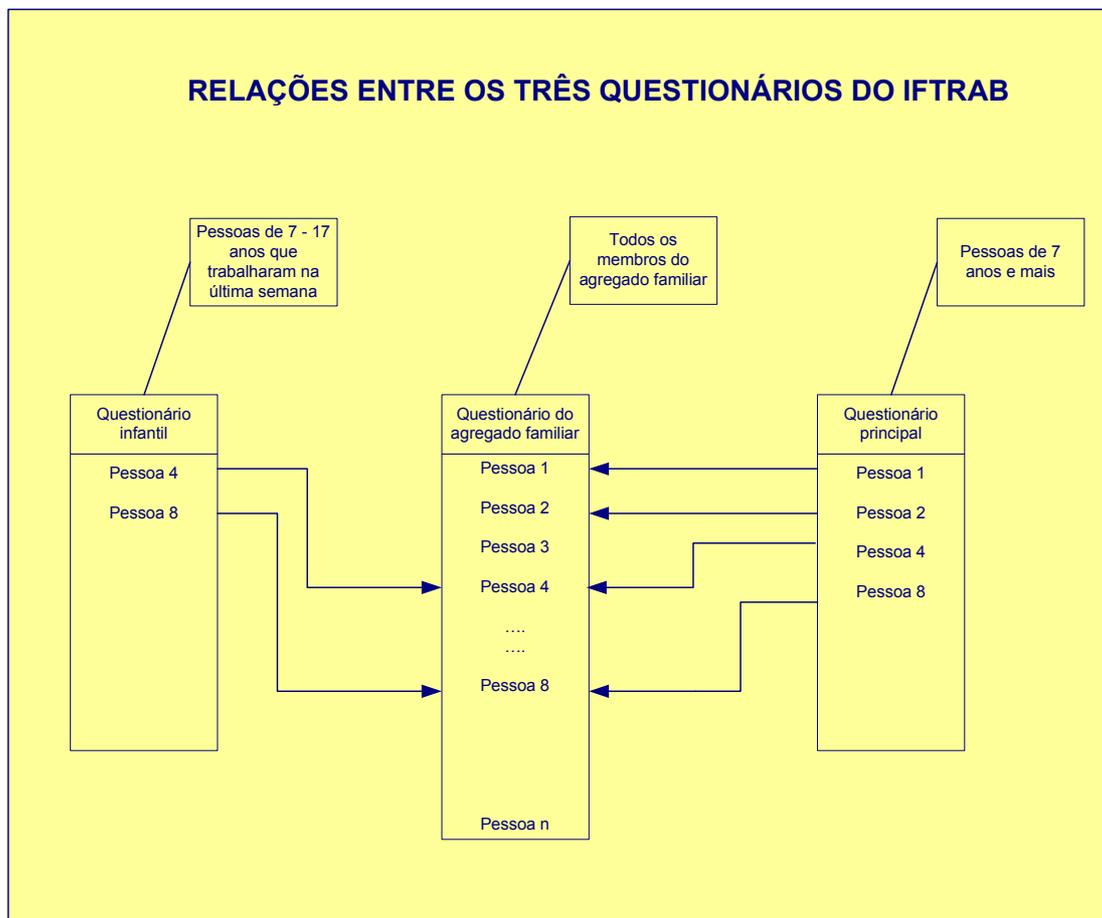
Do ponto de vista do consultor, este objectivo foi atingido, apesar de este constituir apenas um aspecto do processamento estatístico de dados, restando outros problemas de igual ou superior importância que estes, os quais devem ser tomados em conta durante a planificação das operações de um inquérito ou censo. Assim, apesar de o consultor ter dispendido a maior parte do tempo concentrado em cobrir e exercitar conceitos avançados de CSPro, usando para tal exercícios práticos com dados do próprio inquérito, algum tempo foi também dedicado a outros aspectos relevantes do processamento estatístico de dados. Infelizmente, devido ao escasso tempo disponível, apenas uma introdução superficial foi dada.

Basicamente, qualquer operação de inquérito ou censo deve ser controlada por um sistema que forneça, a directores e gestores encarregues pela operação, informação e implicitamente credibilidade em relação à informação obtida, digitada e gravada em suporte magnético. Dando seguimento a este objectivo primário, deveríamos identificar a área primária ou pequena unidade geográfica para acompanhar todas as etapas pelas quais os dados estatísticos passarão. Embora limitemos a abrangência desta descrição apenas às etapas do processamento de dados, a mesma estrutura poderia servir também para outras etapas que antecedem o processamento de dados tais como etapas administrativas. A unidade que melhor funciona e reflecte os nossos propósitos é a Unidade Primária de Amostragem (UPA), pois a mesma foi definida pelo amostrista em termos de localização geográfica, o número de agregados familiares esperados ou unidades de análise esperadas para cada UPA, etc. Adicionalmente, o tamanho desta unidade parece apropriada para (i) o volume de trabalho que os digitadores poderão realizar diariamente e (ii) é a menor área para a qual possuímos informação fidedigna proveniente do desenho amostral, facilitando deste modo a criação do ficheiro de controlo e o acompanhamento do mesmo. Para uma revisão completa da metodologia sugerida, queira por favor ver o anexo 2: “Controlo do inquérito e acompanhamento de UPA’s.

Dentre os vastos conceitos de CSPro passados em revista neste curso, os mais importantes são os seguintes:

1. Relações entre objectos de CSPro. CSPro, mais ou menos à semelhança de bases de dados relacionais, permite definir ligações entre duas tabelas (Chamadas grupos em CSPro). Estas ligações ou relações facilitam a referência entre itens pertencentes a grupos relacionados entre si, eliminando a necessidade de direccionamentos que nem sempre são fáceis de entender. No caso particular do INE, o uso de relações tornava-se necessário devido ao facto de o inquérito à força de trabalho possuir três questionários. O gráfico que a seguir se apresenta, ilustra a situação actual do IFTRAB. O questionário do agregado familiar inclui todos os membros do agregado familiar enquanto os restantes questionários incluem sub-conjuntos dos indivíduos existentes: o questionário principal inclui todas as pessoas de 7 e mais anos de idade e o questionário infantil inclui somente pessoas entre os 7 e os 17 anos de idade

que tenham trabalhado na última semana antes do início do inquérito. Em ambos os casos, a ligação é com o questionário do agregado familiar (através do número do membro) e conseqüentemente, para referenciar o questionário a partir do questionário infantil não existia ligação directa. Contudo, havendo uma ligação comum com o questionário do agregado familiar tornou-se possível estabelecer a ligação entre os questionários infantil e principal.



Conforme ilustrado na figura acima, as relações ou conceitos por detrás das relações constituíam uma necessidade básica para controlar as perguntas de filtro de cada questionário – as quais em ambos os casos envolviam variáveis de diferentes questionários – mas também para controlos de consistência entre variáveis de módulos diferentes.

2. Controlo de lotes. Parece ter havido um mal-entendido nos termos de referência sobre o significado do termo “batch control”, visto o mesmo estar associado à operação de lotes corridos durante o controlo de consistência. O erro resulta porém do facto de o mesmo ter surgido do tempo em que não era possível correr em CSPro outras aplicações e por conseguinte estas aplicações eram corridas no ambiente DOS no modo batch. A partir da altura em que se tornou possível executar, em CSPro, esse tipo de aplicações assim como qualquer outra de ambiente Windows, o modo batch de DOS foi substituído por uma aplicação mais simples de CSPro, a qual, de forma mais rápida, facilita a criação dinâmica de menus e parâmetros e passa-os directamente para aplicações de CSPro. Juntamente com esta facilidade, outros comandos

novos e funções de sistema foram desenvolvidos para simplificar a interacção entre o programador e a geração dinâmica de menus. O resultado foi uma importante descoberta neste processo de encobrir do utilizador final a interface menos amigável de CSPro substituindo-a por uma mais amigável. Esta interface é orientada para um conjunto de aplicações de menus que não permite aos utilizadores finais cometerem erros na administração ou interpretação incorrecta das várias aplicações.

Foi efectuada uma revisão intensiva das técnicas para a criação dinâmica de aplicações orientadas por menus. Exemplos de aplicações para ambos inquéritos foram desenvolvidos e testados à medida que os conceitos eram discutidos, não deixando sombra de dúvidas para mal-entendidos ou dúvidas.

3. Controlos de consistência. De acordo com os termos de referência, este era um tópico muito importante para ser analisado e discutido. Contudo, é importante salientar que as técnicas de CSPro para a implementação de controlos de consistência é um processo simples, uma vez tendo o programador da aplicação uma ideia clara das várias regras de consistência que deverão ser aplicadas nos dados do inquérito. Noutras palavras, o programador da aplicação pode até ser um especialista em CSPro mas se não estiver devidamente claro do que deve ser controlado, não poderá haver um processo adequado de controlo de consistência.

A fonte ou origem dos erros deve ser tomada em consideração no momento de planeamento do processo de validação e consistência dos dados. Erros originados pelos digitadores podem ser eliminados ou pelo menos minimizados no processo de verificação. Estes podem também ser drasticamente reduzidos durante o processo on-line de controlo de consistência assim como pela geração adequada de mensagens de erros. Contudo, erros originados durante a entrevista – pelo inquiridor ou pelo respondente – só podem ser detectados através de um forte conjunto de regras de consistência, o qual requer um profundo conhecimento do(s) questionário(s). Erros originados no campo/entrevista nunca devem ser corrigidos pelos digitadores; as suas capacidades de correcção devem ser rigorosamente limitadas aos seus próprios erros, deixando a correcção dos restantes erros para pessoal especialista mais qualificado. Tomando como base estas premissas haverá duas instâncias para a detecção de erros: (i) o controlo on-line de erros durante a entrada de dados, onde os digitadores corrigem apenas erros originados durante a entrada de dados e (ii) o mesmo processo de correcção é executado após a entrada de dados, também conhecido como processo secundário de correcção em modo batch, o qual tem por objectivo produzir uma lista de erros para áreas pequenas (UPA's) para ser analisada por supervisores e/ou especialistas na matéria, responsáveis estes encargos de corrigir as inconsistências verificadas. Os erros apresentados nessas listas devem ser restringidos àqueles originados no campo, uma vez que presumivelmente os originados pelos digitadores já devem ter sido corrigidos.

O último aspecto abordado levanta outro importante tipo de problema que tem a ver com o tipo de recolha de dados/digitação que se pretende implementar. Apesar de a entrada de dados centralizada ser mais simples de implementar, a correcção de erros originados durante a entrevista é muito mais complexa e pode até distorcer alguns dados. Por outro lado, a entrada distribuída de dados significa um sistema mais

sofisticado de preparação das operações do inquérito, requerendo uma infra-estrutura diferente e por conseguinte a substituição de máquinas normais de registo de dados (computadores) por computadores portáteis e mais caros. O ganho aparece em forma de melhor qualidade de dados, uma vez que novas visitas aos agregados familiares são possíveis quando a quantidade de erros ou a importância dos mesmos assim o justifiquem. A última refinação no processo de captação de dados é a operação “CAPI”, uma vez que toda a entrevista é praticamente levada a cabo pela aplicação CAPI. Isto significa que todos os erros detectados pela aplicação CAPI podem ser corrigidos imediatamente, com a cooperação directa do respondente. Teoricamente, uma vez concluída a entrevista, os dados estão livres de erros e não requerem nenhuma outra aplicação, no futuro, para a detecção de novos erros. O preço a pagar por uma operação CAPI é relativamente elevado, uma vez que é necessário um computador portátil por inquiridor. Adicionalmente, a preparação da aplicação CAPI requer um alto nível de conhecimento e sofisticação. É conveniente mencionar que a SERPRO, com um sub-contrato da Agência Americana para o Desenvolvimento Internacional (USAID) está presentemente realizando um estudo de viabilidade para a implementação de CSPro (a componente interactiva do sistema que corre em “run time”) em computadores de bolso. O software deverá ter a mesma potência e características da versão para computador e ao mesmo tempo fazer uso das vantagens do hardware de computadores de bolso: (mais de 8 horas de duração da bateria, resistência à poeira, baterias rapidamente recarregáveis, baixo peso, ecrã de contacto, conectividade sem fio embutida e baixo preço). O nosso estudo de viabilidade deverá estar concluído na primeira semana de Agosto e uma cópia do referido estudo será enviada ao INE. A conversão do software deverá estar pronta para testes beta no princípio do próximo ano.

4. Exportação de dados para SPSS, SAS e STATA. A nova interface gráfica de exportação de dados foi mostrada e revista em detalhe. Exemplos de exportação de dados num registo de variáveis com base nos dados dos três questionários do IFTRAB foram desenvolvidos usando a opção RELATIONS de CSPro. De forma geral, a exportação de dados para qualquer um dos sistemas de análise estatística não deve constituir nenhum problema quando se usa a nova interface gráfica de CSPro.
5. Tabulações usando a corrente interface gráfica. Apesar de a presente versão de tabulações cruzadas de CSPro ser ainda básica, dois objectivos foram atingidos com a sua revisão e análise.: (i) a fácil criação de tabelas simples e (ii) a introdução da nova e poderosa interface gráfica de tabulação cruzada de CSPro, a qual deverá ser lançada em Setembro deste ano. É importante salientar que o novo módulo de tabulação cruzada tornará possível a criação de tabelas complexas a par com outras medidas estatísticas de tendência central (média, mediana e moda), dispersão (mínimos, máximos, variâncias, desvio padrão, erro padrão) e outras tais como proporções, percentis, percentagens, etc. É altamente recomendável que técnicos do DISI tenham pelo menos 2 semanas de seminário para esta importante componente de CSPro, concluindo assim o seu treinamento em CSPro.

### 3. RECOMENDAÇÕES

Os principais problemas identificados no processamento dos dados do IFTRAB provêm da falta de um plano detalhado para a detecção de diferentes tipos de erros que podem ser encontrados nos dados estatísticos. A produção do referido plano deveria resultar da criação de um grupo multi-disciplinar, do qual deveriam fazer parte técnicos de informática assim como demógrafos, sociólogos e em geral especialistas na matéria de forma a contribuírem para a conceptualização do plano no qual nenhum detalhe deve ser deixado ao acaso.

O plano tem por objectivo revelar e reparar ou corrigir erros a dois diferentes níveis, os quais requerem diferentes estratégias: (i) os problemas que tendem a distorcer o universo em estudo, os quais podem ser afectados pela colocação errada de pequenas áreas em diferentes áreas geográficas/administrativas. Já foi sugerida uma aplicação de controlo com o objectivo de prevenir que este tipo de erros aconteçam e que a mesma deve ser parte do plano de processamento de dados de qualquer inquérito/censo. (ii) uma especificação detalhada de todos os controlos de consistência a serem executados a nível da unidade em estudo. A unidade em estudo deve ser o agregado familiar, a machamba ou qualquer outra unidade a ser analisada, dependendo do tipo de inquérito. Esta lista de especificações requer um conhecimento profundo do questionário assim como das inter-relações entre os diferentes módulos do questionário e por conseguinte, a cooperação de grupo multi-disciplinar é uma contribuição muito valiosa.

As preocupações anteriormente mencionadas têm por objectivo produzir dados de inquéritos/censos que tenham sido rigorosamente e metodologicamente tratados, passo a passo, de forma a prevenir a ocorrência de erros que possam distorcer o resultado estatístico que se esperar obter. O passo seguinte deve ser a produção destas estatísticas em termos de tabulações cruzadas e parâmetros estatísticos ou indicadores que estaticistas e fazedores de políticas realmente necessitam para atingir o objectivo final, o qual deve ser a investigação e análise dos dados obtidos. Para atingir este objectivo final, é altamente desejável que os técnicos de informática obtenham treinamento apropriado no novo módulo de CSPro a ser lançado brevemente. A compilação deste treinamento iria deixar os técnicos do DISI capacitados em todas as áreas do processamento de dados de inquéritos.

## **Apêndice 1: Termos de referência**

### **TERMOS DE REFERÊNCIA**

Missão de curta duração sobre processamento de dados para o Inquérito à Força de Trabalho para o pessoal do DISI de 12 a 26 de Junho, no âmbito da capacitação institucional ao INE para o período 2003 – 2007, com assistência Escandinava.

Consultores: SERPRO S.A.

Contraparte: senhor Tomás Bernardo, director-adjunto da DICRE

#### **1.1 Antecedentes**

O Inquérito à Força de Trabalho, baseado no Questionário Básico de Indicadores de Bem-Estar (QUIBB) e correspondentes módulos rotativos, iniciou em 2000. O Inquérito à Força de Trabalho (IFTRAB), que termina em 2005, é o primeiro ciclo nesta nova abordagem. O trabalho realizado no IFTRAB mostra que o Instituto Nacional de Estatística (INE) está ganhando competência na área de planeamento e logística mas também demonstra fraquezas no planeamento e execução de operações relacionadas com o processamento de dados. Por conseguinte, o INE necessita de reforçar a sua competência no processamento de dados. O Inquérito Demográfico e de Saúde (IDS) de 2003 sofreu maioritariamente dos mesmos problemas e com o aumento do nível de conhecimentos, espera-se que estes problemas sejam evitados no futuro. Os presentes termos de referência propõem a realização de uma missão com o objectivo de rectificar esta situação, através da realização de uma acção de formação em CSPro para o pessoal envolvido no processamento de dados.

Para além da relevância imediata para o IFTRAB, esta missão deverá ajudar na execução do Inquérito ao Sector Informal (INFOR) a ter lugar no corrente ano de 2005 e financiado com fundos de um projecto italiano, recentemente iniciado. Uma componente deste é o inquérito aos agregados familiares, o qual poderá ser incorporado ao IFTRAB, pelo menos com algumas perguntas básicas. A outra componente inclui actividades relacionadas com a preparação do censo da população 2007.

#### **1.2 Principais razões para a realização da missão**

O INE carece de experiência profunda no processamento de dados usando CSPro, no contexto do IFTRAB. Assistência é, portanto, necessária para a capacitação nesta área.

#### **1.3 Beneficiários da missão**

A missão será benéfica para o INE assim como para usuários de dados estatísticos sobre força de trabalho e emprego.

#### **1.4 Objectivos da missão**

O objectivo geral da missão é prestar assistência e capacitar o INE no uso de CSPro, em particular o controlo de lotes e de consistência. A missão deve de igual modo identificar fraquezas no processamento de dados e, caso necessário, propor mudanças nos procedimentos e/ou futuras necessidades de formação.

#### **1.5 Resultados esperados**

- Pessoal local treinado no uso avançado de CSPro.
- Um documento com metodologias descrevendo procedimentos para o processamento de dados usando CSPro.
- Problemas no processamento de dados do IFTRAB 2004/5 resolvidos.

#### **1.6 Agenda da missão**

A agenda da missão será especificada no primeiro dia da reunião em Maputo

#### **1.7 Termos de referência**

- Treinamento de pessoal informático em programação avançada em CSPro, nomeadamente em:
- Controlo de lotes (ligação de diferentes tipos de questionários)
- Controlo de consistência para respostas múltiplas

#### **1.8 Actividades a realizar pelo INE para facilitar a missão**

- Fornecer ao consultor informação sobre problemas identificados no processamento de dados do IFTRAB
- Assegurar a disponibilidade de pessoal do INE envolvido no processamento de dados do IFTRAB
- Preparar e fornecer ao consultor documentos relevantes e informação tal como documentação sobre como o processo de entrada de dados do IFTRAB deve ser levado a cabo.
- Criar boas condições de trabalho para o consultor

#### **1.9 Consultor e contraparte**

Consultores: SERPRO

Principal contraparte: senhor Tomás Bernardo, director-adjunto para a área de Informática no INE.

### **1.10 Duração da missão**

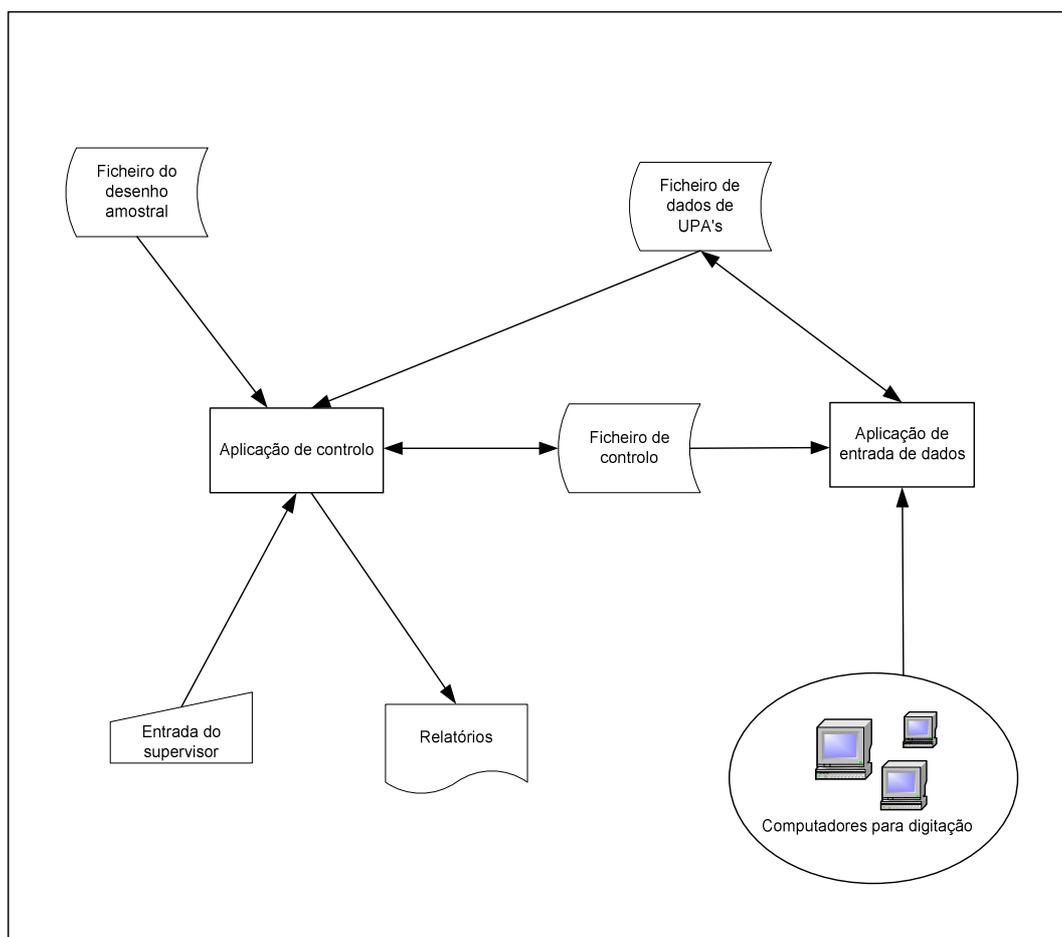
12 A 26 de Junho.

### **1.11 Relatório**

O(s) consultor(es) deverá preparar o relatório rascunho da missão a ser submetido ao INE. A estrutura do mesmo deverá ser no formato da DANIDA (vide artigo 3 do contrato). O conteúdo principal deverá contemplar as metodologias e procedimentos do processamento de dados.

A contraparte deverá garantir que a versão final impressa do relatório inclua uma versão compreensível em português, caso o relatório final esteja em Inglês/Espanhol e, pelo menos, um resumo em Inglês, se o relatório principal estiver em Português.

## Anexo 2: Controlo de inquérito e acompanhamento de UPA's



Na base deste tópico, é descrita a metodologia para correctamente representar o universo em estudo. Identificamos a UPA como o conjunto de dados micro ideal a ser monitorado e acompanhado, cujo diagrama acima apresentado descreve os diferentes elementos que ilustram a nossa metodologia.

Conforme mostra o diagrama, existem duas aplicações diferentes ligadas por um ficheiro de dados comum (ficheiro de controlo), o qual executa actividades complementares e que a seguir se descrevem:

### **Aplicação de controlo**

Um erro comum no processamento de dados estatísticos é a má interpretação do universo em estudo, seja por duplicação de pequenas áreas geográficas ou administrativas ou simplesmente por omissão das mesmas. Em qualquer dos casos, os erros resultantes desta situação podem ter um impacto importante nos resultados obtidos, uma vez que pode-se estar sobrestimando a população de algumas áreas ou subestimando os dados no segundo caso. A prevenção contra a má interpretação do universo em estudo é um processo relativamente simples mas é necessário obedecer a um plano organizado, tal como o que a seguir se apresenta.

Esta aplicação deve realizar, pelo menos, as seguintes actividades:

Criação do ficheiro de controlo. Usando a informação básica proveniente do desenho da amostra, a aplicação gerará o ficheiro de controlo. O ficheiro de controlo terá um registo por UPA, usando as áreas geográficas/administrativas (por exemplo: província, distrito, etc.) às quais a UPA pertence como parte da chave ou identificação directa. Com base no ficheiro de dados do desenho amostral, a seguinte informação deverá ser gerada automaticamente pela aplicação:

- Identificação completa da UPA;
- Número esperado de unidades de análise pela UPA;
- Peso ou factor de expansão;
- Outra informação relevante;

Para além da informação anteriormente apresentada, no momento da criação do ficheiro de controlo da aplicação, todos os restantes componentes do registo deverá ser inicializado para os seus valores iniciais por defeito. Assim, esta é uma actividade que deve ser executada no começo da operação mas nunca depois do ficheiro ter sido actualizado com informação actual.

Recepção da UPA. Após a recepção de informação do campo, a colecção de questionários (UPA) é enviada de volta aos serviços centrais (caso a entrada de dados for centralizada) ou para a Delegação Provincial (se a entrada de dados for distribuída). A captura de dados por CAPI (entrevista pessoal por computador) requer um acompanhamento ligeiramente diferente, o qual não será aqui descrito. Os itens de dados que podem ser capturados durante a recepção da UPA devem ser:

- Data de recepção
- Contagem manual do número de unidades (agregados familiares ou outra unidade) para a UPA

A data de recepção, assim como a contagem manual de unidades permitirá a elaboração de relatórios sobre UPA's em falta para determinadas áreas assim como um relatório avançado orientado aos administradores da operação de inquérito.

- Atribuição da UPA ao digitador primário: o supervisor de entrada de dados deve atribuir cada UPA a um digitador específico. A atribuição é reconhecida pela captura da seguinte informação:
  - Data da digitação da UPA pelo digitador primário;
  - Código de identificação do digitador;
  - Associado à informação descrita anteriormente deverá ser também reconhecida a data em que o digitador concluiu de digitar a UPA, assim como o número de unidades (agregados familiares ou outra unidade) digitados, apesar de a data destas operações diferir da obtida e gravada no ficheiro de controlo.
  - Data em que a digitação primária terminou;
  - Número de unidades digitadas (contadas directa pela aplicação);

- Atribuição da UPA para a digitação secundária: Apesar de CSPro ser capaz de fazer a verificação de dados “on-line”, significando que os dados digitados estão em “background” e os dados para verificação estejam em “foreground”, nós sugerimos que os dados sejam digitados em dois ficheiros diferentes e comparados mais tarde com recurso a um utilitário de CSPro. A razão de ser deste procedimento é a obtenção de melhor estatística de erros, permitindo desse modo ao supervisor ajuizar melhor a qualidade dos dados digitados pelo digitador. A informação derivada desta atribuição é similar à da digitação primária. Note-se que ambas atribuições podem ser feitas em paralelo, visto os ficheiros serem independentes. A seguinte informação deve ser registada:
  - Data em que a atribuição da digitação secundária teve lugar;
  - Código de identificação do digitador;
  - Data em que a digitação secundária da UPA terminou;
  - Número de unidades digitadas (contadas directamente pela aplicação);
- Comparação de item por item do ficheiro de dados primário e secundário guiada por um dicionário comum: Esta operação é efectuada automaticamente por um utilitário de CSPro e produz um relatório das diferenças encontradas. O programador da aplicação pode incluir/excluir variáveis específicas do processo de comparação ou simplesmente incluir todas elas. O relatório é muito importante para avaliar a qualidade dos digitadores e permitir a detecção atempada daqueles que, potencialmente, podem perigar a operação. Uma vez demonstrada a igualdade destes ficheiros para a digitação primária e secundária, a digitação de dados é considerada concluída através da inclusão da data em que a UPA digitada foi aprovada.

Tendo o ficheiro de controlo de dados sido actualizado e a aplicação de entrada de dados sido ajustada para trabalhar harmoniosamente com o ficheiro de controlo, podemos evitar (i) duplicação de UPA’s, uma vez que as mesmas terão sido atribuídas especificamente pelo supervisor de entrada de dados a um digitador específico; (ii) a omissão de UPA’s, uma vez que o sistema de relatórios facultado pela aplicação de controlo indicará claramente quais UPA’s estão em falta a qualquer altura, durante a operação do inquérito.

## **Annex: Relatório do curso de CSPro**

O INE iniciou, em 2000, um sistema integrado de inquéritos aos agregados familiares. O mesmo é composto por módulos rotativos, sendo um desses módulos o Inquérito à força de Trabalho (IFTRAB), iniciado em 2004 e estando previsto o seu término no presente ano.

Um dos problemas identificados neste inquérito relaciona-se com o processamento de dados do mesmo com recurso ao software CSPro, nomeadamente a ligação de questionários assim como o tratamento de respostas múltiplas. Estes problemas resultavam do facto de os técnicos do DISI envolvidos no processamento de dados não estarem devidamente capacitados a solucionar este tipo de problemas.

Com o objectivo de ministrar um curso avançado de programação em CSPro, o INE solicitou uma consultoria à Serpro, uma empresa da República do Chile envolvida juntamente com o US Bureau of Census no desenvolvimento do referido software.

De 13 a 24 de Junho teve lugar o curso avançado de programação em CSPro no qual participaram Tomás Bernardo, Eugénio Matavel e Bruno Couto do INE e 4 técnicos do Ministério da Agricultura (MINAG).

O curso foi ministrado em forma de “ on the Job training” envolvendo uma parte teórica e outra de exercícios tomando como base o IFTRAB e o TIA (Trabalho de Inquérito Agrícola) do MINAG. Para além dos aspectos indicados nos termos de referência para o curso, o consultor capacitou os técnicos de conhecimentos adicionais para o processamento de dados de inquéritos de índole similar ou mais complicados tais como criação de menus para as diferentes etapas do processamento de dados, controlo de rendimento de digitadores, controlo de áreas geográficas com recurso a ficheiros externos, produção de tabelas cruzadas entre outros.

Antes do fim do curso foi organizado um encontro no qual participaram técnicos da Demovis, envolvidos na análise de dados. A finalidade do referido encontro era fazer uma demonstração sobre diferentes tipos de consistência implementados, os quais foram considerados de bons e adequados ao processamento dos dados.

No fim da missão do consultor foi realizada uma reunião de avaliação dos objectivos do curso. Participaram no referido encontro

Fátima Zacarias, directora da Demovis,  
Arão Balate, director da DCI,  
Tomás Bernardo, director-adjunto da DICRE,  
Júlio Ortuzar, consultor da empresa Serpro,  
Hans-Erik Altvall, team leader do programa Escandinavo  
Dag Roll-Hansen, consultor para o pelouro demográfico

O INE fez uma avaliação positiva do curso, pois o mesmo conseguiu atingir os objectivos traçados e os problemas identificados no processamento de dados foram resolvidos. Por sua vez o consultor Ortuzar considerou que os participantes do curso assimilaram as matérias leccionadas e estão dotados de ferramentas necessárias para a

realizar com sucesso a tarefa de processamento de dados do IFTRAB e outro tipo de inquéritos natureza similar.

### **Recomendações**

O DISI deveria, internamente, organizar um curso básico de programação em CSPro para os restantes técnicos do departamento, de forma a evitar depender de apenas 2 técnicos com domínio deste software.

O DISI juntamente com as áreas de produção estatística deve elaborar manuais de validação e consistência de dados para evitar que o processamento de dados (consistência) seja feita de forma “ad-hoc”. Por outro lado, os referidos manuais serviriam de documentação para consulta e para a realização de outros inquéritos.

O software CSPro ainda está em fase de desenvolvimento, faltando ainda o módulo de criação de tabelas complexas. Quando o módulo em falta estiver concluído será possível programar todas as tabelas dos planos de tabulação, facilitando desta forma o trabalho do grupo de análise dos dados. De momento, os dados processados são exportados para SPSS para a sua análise. Uma acção de formação para o processamento de tabelas complexas torna-se necessária para o grupo que tomou parte neste curso, logo que o módulo em questão esteja concluído.

Elaborado por: Tomás Bernardo