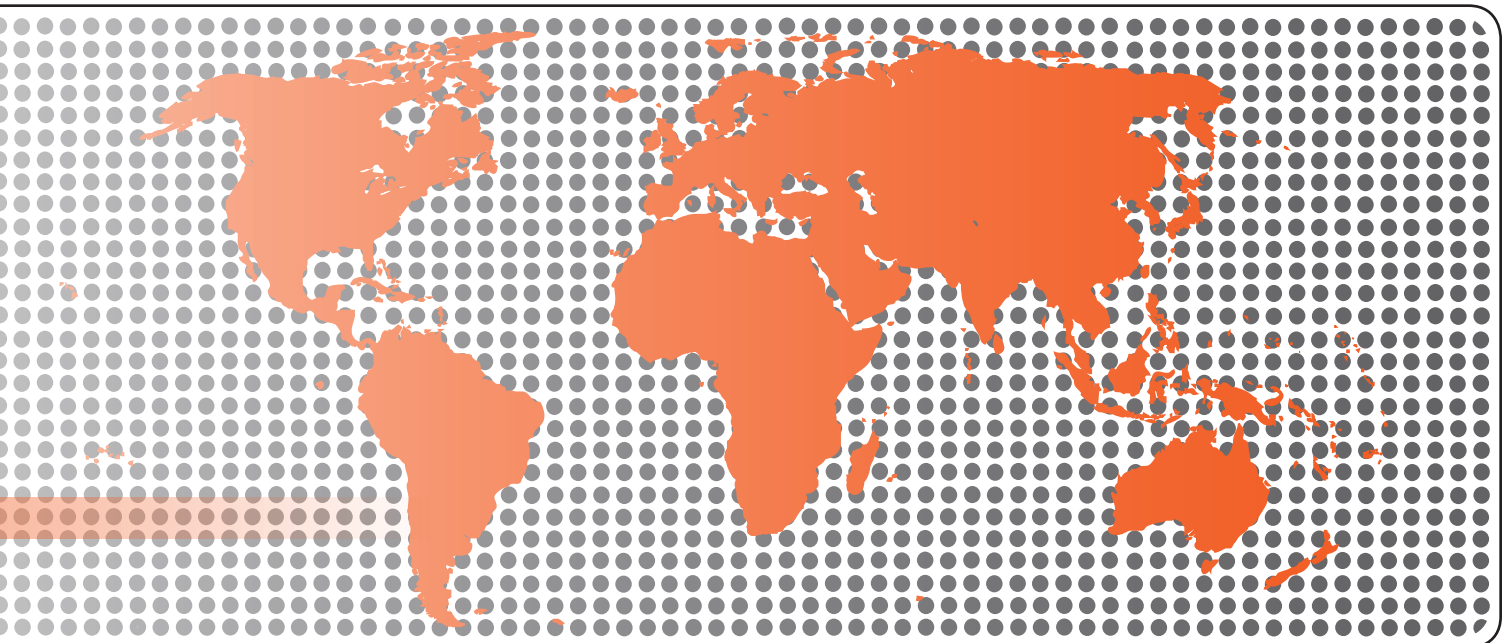GTSS Global Adult Tobacco Survey (GATS)

Data Management Implementation Plan

Global Adult Tobacco Survey (GATS)
Data Management Implementation Plan

Version 2.0
November 2010

Global Adult Tobacco Survey (GATS)
Comprehensive Standard Protocol

....

GATS Questionnaire
Core Questionnaire with Optional Questions
Question by Question Specifications

GATS Sample Design
Sample Design Manual
Sample Weights Manual

GATS Fieldwork Implementation
Field Interviewer Manual
Field Supervisor Manual
Mapping and Listing Manual

GATS Data Management
Programmer's Guide to General Survey System
Core Questionnaire Programming Specifications
Data Management Implementation Plan
Data Management Training Guide

GATS Quality Assurance: Guidelines and Documentation

GATS Analysis and Reporting Package
Fact Sheet Template
Country Report: Tabulation Plan and Guidelines
Indicator Definitions

GATS Data Release and Dissemination
Data Release Policy
Data Dissemination: Guidance for the Initial Release of the Data

**Tobacco Questions for Surveys: A Subset of Key Questions
from the Global Adult Tobacco Survey (GATS)**

Suggested Citation

Acknowledgements

GATS Collaborating Organizations

- Centers for Disease Control and Prevention

- CDC Foundation

- Johns Hopkins Bloomberg School of Public Health

- RTI International

- University of North Carolina Gillings School of Public Health

- World Health Organization

Financial Support

Disclaimer: The views expressed in this manual are not necessarily those of the GATS collaborating organizations.

Contents

Exhibits

1. Introduction

Tobacco use is a major preventable cause of premature death and disease worldwide. Approximately 5.4 million people die each year due to tobacco-related illnesses – a figure expected to increase to more than 8 million a year by 2030. If current trends continue, tobacco use may kill a billion people by the end of this century. It is estimated that more than three quarters of these deaths will be in low- and middle-income countries[1]. An efficient and systematic surveillance mechanism is essential to monitor and manage the epidemic.

The **Global Adult Tobacco Survey** (GATS), a component of Global Tobacco Surveillance System (GTSS), is a global standard for systematically monitoring adult tobacco use and tracking key tobacco control indicators. GATS is a nationally representative household survey of adults 15 years of age or older using a standard core questionnaire, sample design, and data collection and management procedures that were reviewed and approved by international experts. GATS is intended to enhance the capacity of countries to design, implement and evaluate tobacco control interventions.

In order to maximize the efficiency of the data collected from GATS, a series of manuals has been created. These manuals are designed to provide countries with standard requirements as well as several recommendations on the design and implementation of the survey in every step of the GATS process. They are also designed to offer guidance on how a particular country might adjust features of the GATS protocol in order to maximize the utility of the data within the country. In order to maintain consistency and comparability across countries, following the standard protocol is strongly encouraged.

> GATS manuals provide systematic guidance on the design and implementation of the survey.

1.1 Overview of the Global Adult Tobacco Survey

GATS is designed to produce national and sub-national estimates among adults across countries. The target population includes all non-institutionalized men and women 15 years of age or older who consider the country to be their usual place of residence. All members of the target population will be sampled from the household that is their usual place of residence.

GATS uses a geographically clustered multistage sampling methodology to identify the specific households that Field Interviewers will contact. First, a country is divided into Primary Sampling Units, segments within these Primary Sampling Units, and households within the segments. Then, a random sample of households is selected to participate in GATS.

The GATS interview consists of two parts: the *Household Questionnaire* and the *Individual Questionnaire*. The *Household Questionnaire* (household screening) and the *Individual Questionnaire* (individual interview) will be conducted using an electronic data collection device.

> The GATS interview is composed of two parts: *Household Questionnaire* and *Individual Questionnaire*. These questionnaires are administered using an electronic data collection device.

---

[1] Mathers, C.D., and Loncar D. Projections of Global Mortality and Burden of Disease from 2002 to 2030. *PLoS Medicine*, 2006, 3(11):e442.

At each address in the sample, Field Interviewers will administer the *Household Questionnaire* to one adult who resides in the household. The purposes of the *Household Questionnaire* are to determine if the selected household meets GATS eligibility requirements and to make a list, or roster, of all eligible members of the household. Once a roster of eligible residents of the household is completed, one individual will be randomly selected to complete the *Individual Questionnaire*. The *Individual Questionnaire* asks questions about background characteristics; tobacco smoking; smokeless tobacco; cessation; secondhand smoke; economics; media; and knowledge, attitudes, and perceptions about tobacco.

1.2   Use of this Manual

The purpose of the ***Data Management Implementation Plan*** (DMIP) is to provide a description of the procedures, practices, and resource information for GATS data management activities inclusive of data extraction, format, transfer, and chain of custody from the interviewer to the global level. The plan was developed in accordance with the GATS task deliverables document and is maintained by the GATS Data Coordinating Center (DCC).

The goal of the DMIP is to provide vital planning and implementation information to the primary GATS partners and to the GATS implementing agencies that reside in each country. It addresses the roles and responsibilities of the GATS partners and implementing agencies and is targeted primarily for use by the following audiences:

- the country staff/institution responsible for planning, acquiring, and managing IT and human resources to carry out GATS and

- the country IT staff or data managers who are responsible for maintaining, transferring, consolidating, and auditing GATS data collection at the interviewer, primary sampling unit (PSU), Regional and/or Country level.

This document also provides specific data processing and management activities, and procedures and recommendations to be performed in the course of data collection and consolidation. It should be noted that whenever possible existing GATS resource documents (listed in **Section 1.4** below) are referenced because explicit instructions and specifications are found in those materials. **Appendix A** provides a glossary of terms used in the document.

1.3   Structure of the Data Management Implementation Plan

Regardless of the data collection method adopted by each country, data collection, consolidation, and quality control (QC) are performed at four main levels: (1) Field Interviewer, (2) PSU/Regional, (3) Country, and (4) Global (DCC). **Exhibit 1-1** describes the data collection workflow for each of these levels. Some countries may opt to bypass the PSU/Regional level and move data from the Field Interviewers' iPAQs[2] to the country database directly.

---

[2] GATS uses General Survey System (GSS) software, which is designed to run on a Windows Mobile platform and has been tested and implemented using a Hewlett-Packard (HP) iPAQ© handheld PDA computer. Use of "iPAQ" is for identification only and does not imply endorsement by any of the GATS collaborating organizations.

Exhibit 1-1.   GATS Data Collection Workflow

Facilities	Suggested Resource	Main Deliverables
Global—Data Coordinating Center (DCC)	1 system/database administrator	Provide training to country staff
	1 technical support	Provide remote technical assistance
	1–2 GATS trainers	Collect, integrate, and manage country data and reports
		Perform QC
		Provide quality assurance and analytic support
Country Database	1 system/database administrator	Provide equipment for country
	1 country technical support	Design local questionnaire
	2–4 survey and IT trainers	Load questionnaire on iPAQs
		Provide training and support
		Consolidate and manage data
		Perform QC
		Analyze and report country data
		Coordinate data release, support, and issues with DCC
Primary Sampling Unit (PSU) Database	1 data manager per PSU	Schedule and assign sample to Field Interviewers
		Collect and aggregate data from iPAQs
		Perform QC
		Upload data to country database
Handheld computer—iPAQs	Field Interviewers	Attend GATS field training
		Collect data
		Upload data from iPAQs to PSU database

Resource requirements, including hardware, software, and personnel, along with functional responsibilities and deliverables at each level, are explained in the sections below.

First, however, readers should familiarize themselves with the following list of technical resources, which provide in-depth details of the GATS questionnaire and the recommended electronic data collection technology, and most importantly with the Data Governance guidelines to ensure clear understanding and delineation of the responsibilities and expectations between and among the multitude of players.

1.4   Technical Resources

A series of manuals and technical documents have been created for the countries to follow as they design sampling plans, develop data collection protocols, and collect data for GATS. These manuals are designed to provide the countries with the minimum requirements issued by CDC and WHO on different aspects of the GATS sample design and data collection and management methodologies. The manuals are also designed to provide guidance on how a country might adjust features of the GATS protocol to maximize the utility of the data for researchers within the country. Considerable flexibility is being offered to countries in the GATS protocol; however, it is important for each country to follow the features that are

required by CDC and WHO. These requirements are being issued to maintain consistency in the implementation of GATS between countries. Maintaining this standardization is imperative because it will provide consistency and allow for data integration and analytic comparisons over countries.

Questionnaire Development

GATS Programmer's Guide to General Survey System—provides information on the GSS set of programs and applications. Information in the document includes the technical and business architecture along with the requirements and usage for the system. This document is written for the IT staff that will be developing and supporting GSS activities.

GATS Core Questionnaire Programming Specifications—contains specifications at the field/variable level for each field in the core instrument.

Country GATS Questionnaire—the survey instrument to be implemented at the Country level. Each country will have a country-specific Q x Q annotated questionnaire that reflects changes they have applied to the core, such as variables that were added or dropped.

Field Implementation

GATS Field Interviewer Manual—provides step-by-step guidance to Field Interviewers.

GATS Field Supervisor Manual—presents requirements and guidelines for the GATS Field Supervisors who supervise Field Interviewers in countries that are using iPAQs.

Data Management

GATS Data Management and iPAQ PowerPoint Presentations—used at various meetings, including technical site visits, the presentation spans GATS data management aspects including presite visit preparations and pretest activities.

Checklist for Evaluating Third-Party Survey Systems—contains specifications used to assess the country-specific (non-GSS) survey system.

Country IT Survey—gathers country-specific requirements information relating to survey operations, hardware, software, and telecommunication so that a data management model can be selected and adapted.

Data Release Policy—describes requirements for preparing data files for the DCC and wider release.

Technical Support

Presite Visit Checklist—a list of activities to be performed by both the country and technical support teams 4 to 6 weeks prior to the pretest site visit.

GATS Technical Support Plan—a brief document that outlines the level of support and resources for the GATS country pretest and main study technical assistance.

1.5   Data Governance

Data governance refers to the rules and regulations for managing the availability, usability, integrity, and security of data as an asset. GATS data governance includes a structure that defines the policies and procedures and a plan to implement them. To accomplish the goals of GATS, it is necessary to put personnel, policies, procedures, and organizational structures in place to make data accurate, consistent, secure, and available. GATS data governance is planned, managed, and implemented through a multinational structure as described in detail below.

To aid in the establishment of GATS data governance, two documents have been created. The *GATS Data Management Implementation Plan* (DMIP) exists to guide the standardization of Country-level GTSS GATS data management activities and to serve as a reference for planning and implementation. In conjunction with the DMIP, the *GATS Data Release Policy* details the policies regarding the release of the GTSS GATS data to users.

Data Management Planning

Development and Approval of Data Policies, Standards, and Procedures

All participants in this structure are stakeholders in the data management policies, procedures, and recommendations process. The DCC drafts and finalizes the policies and plans. The GATS Partners and the GTSS Advisory Committee (GTSAG) review and approve plans and policies. These approved plans and policies are subsequently made available to participating GATS country implementing agencies for their feedback and implementation. The DCC will maintain and release updated versions as plans and policies change.

Data Stewardship

Data stewardship is the identification of positions in organizations that are accountable for the definition, production, and use of the data. Although the implementing agency is responsible for identifying and providing staff resources, this plan provides guidance in this effort. **Appendix B** provides a potential list of staff that may be used to support and perform GATS data management activities, along with their job descriptions. For countries needing to acquire IT staff for the implementation of GATS, this resource may serve to help evaluate the people on their ability to be responsible for the data. Although in-country staffing is expected to be composed of numerous medical, technical, and administrative positions residing at Regional and national levels, this plan solely addresses technical staff and their needed resources. Although various technical position titles may vary, staffing may include one to three individuals in the following roles, depending on the number of field staff and the technical aspects required in the country: IT/Computer Specialist, Data Manager, Database Administrator, Statistical Programmer, and Applications Programmer.

Data Management Control

Coordinating Data Activities

Once planned, data management activities must be coordinated. This plan covers GATS procedures for performing these activities specific to the following:

- planning, supervision, training, and control over data management

- data architecture and design

- defining, monitoring, and improving data quality

- Global-level concerns

  – *database design, implementation, and support*

  – *reviewing data and performing QC activities*

  – *receiving and documenting data files*

- enabling reporting (e.g., integrating, controlling, and providing metadata; creating analysis datasets at the international level)

Monitoring and Managing GATS Data

Within the numerous GATS data management activities, data monitoring will require significant ongoing interaction between the GATS partners and the implementing agencies. The goal of GATS is to provide guidance for a sustainable in-country surveillance system as well as an integrated global-level data repository that allows for user analysis. Survey monitoring (e.g., field progress, data collection, transmission, and global acquisition) is a key coordinated, ongoing activity that will require all stakeholders to be involved.

Communication

Communication is extremely important for successful data governance. The GATS partners must tell implementing agencies what steps are being taken and why and inform relevant communities to incorporate their ideas and feedback. Input and feedback make governance efforts more effective in achieving goals and are vital to the success of GATS. Communication mechanisms include:

- ongoing technical assistance,

- ongoing GATS partners meetings, and

- partner conferences.

2. General Data Management Planning Guide

Each participating GATS country will submit a GATS survey proposal that include the details for data collection and data management plans, including technical and human resource requirements. The proposal will include the final determination of data collection technology, if other than that recommended and provided by GATS. The final proposal must include description of the following:

- determination and description of the number of PSUs or Regional survey data collection centers

- number of interviewers and other technical and data management staff at the National Data Center and at each PSU

- description of the interview data collection technology (hardware, software, etc.), if other than the hardware and software recommended by GATS

- description of data consolidation and management technologies at the PSU and Country level, including computers, databases, Internet capabilities, and data quality analysis software

- sampling case assignments and management procedures, including the tracking of follow-up and incomplete cases

- procedures for data gathering and consolidation from interviewers to PSU/Regional databases

- data quality testing and remediation procedures at the PSU/Regional level

- data transfer, aggregation, and chain-of-custody procedures from PSUs/Regions to the country central database

- procedures for transfer of data approved for release from the country to the DCC

- technical and data management training plans for Field Interviewers, supervisors, and IT staff at the PSU/Regional/national level

- interviewer training plans

- country technical and data management staff training plan

2-2

3. Field Data Collection

GATS proposes and strongly encourages countries to use handheld electronic data collection technologies (iPAQs) designed specifically for the GATS to ensure timely, cost-effective, and quality-controlled survey data collection and consolidation. The GATS will ensure the delivery of iPAQs, software to develop electronic GATS questionnaires in the country's host language, and procedures to transfer and consolidate data from iPAQs to PSU databases to National Data Centers to the DCC global repository.

Countries with established national electronic data collection infrastructure may choose to deploy their existing technology and expertise for field data collection instead of the handheld technology recommended and funded by the GATS. However, closer collaboration between Countries using their existing infrastructure and the GATS partners will be necessary to ensure that GATS data are collected in accordance with agreed-on specifications and quality.

3.1 Electronic Data Collection Devices (iPAQs)

The HP iPAQ HX2490c was used in early GATS work, and more recently the HP iPAQ Model 210 has been the device of choice for the GATS data collection. The iPAQ 210 device has an external SD slot and is supplied with an extra battery and 2 SD cards. In certain countries, other iPAQ handheld models have been used to meet in-country requirements. The CDC Foundation will procure and deliver the appropriate number of iPAQs to each country in time for GATS survey data collection. Additional units (10%) will be included with each shipment for use as backup units. All iPAQs will remain in the country after completion of the GATS survey. IPAQs are suitable for use in a wide range of applications including non-GATS surveys.

3.2 Electronic Survey Questionnaire Software and Data Storage

Significant efforts went into developing an electronic GATS questionnaire for the iPAQ with appropriate data entry validation, skip patterns, and standard data coding and formatting. The Survey Questionnaire was developed using the GSS software. This software will be made available to GATS host countries. GSS questionnaires are certified for use with the iPAQ only and do not support laptops or tablet PCs. Data collected by the GSS iPAQ questionnaire are stored internally in a Microsoft SQL Compact Edition database format.

3.3 iPAQ and Questionnaire Software Installation and Technical Support

Training on the installation, setup, and operation of the GSS software and iPAQ hardware supplied by GATS will be delivered to "key staff" in each country as part of the pretest implementation plan. "Key staff" will in turn install and set up the iPAQs for the final survey and provide training and technical support to PSU staff and interviewers as part of the final survey implementation plan. A training guide has been prepared to be used for onsite training. This will insure systematic coverage and a consistent approach in the training that takes place in each country. Training and technical support will be available in the following areas:

- setting up the iPAQ systems' hardware

- selecting the most effective data management model for each country

- developing questionnaire forms using GSS

- modifying the core GATS questionnaires

- setting up multiple language versions of the questionnaires

- loading sample files and initializing the case management systems

- training and supporting Field Interviewers and Supervisors

- managing and performing the data management tasks to move data from the iPAQs back to the national data center

The *GATS Programmer's Guide to General Survey System* provides detailed instructions on how to prepare equipment and questionnaires for the field.

Appendix C provides a list of the main setup tasks that are required on site.

The *GATS Data Management Training Guide* will be available to direct training activities. This pretest training and its guide are aimed at training those in-country staff who will eventually train all staff for the production full-scale survey. Technical support will be available after the pretest and during full-scale survey operations. Part of the pretest setup and training will focus on developing a support hierarchy within the host country and communications and support mechanisms to provide second- and third-tier support to in-country staff.

3.4 Data Collection Procedures

The data collection procedures supported by GSS and the operations required by field staff are detailed in the field manuals that have been developed. The *GATS Field Interviewer Manual* provides step-by-step guidance to Field Interviewers. The *GATS Field Supervisor Manual* presents requirements and guidelines for the GATS Field Supervisors who supervise Field Interviewers.

The GSS software and its implementation, as described in the above manuals, will lead to building iPAQ-based databases that contain the questionnaire data collected by each interviewer. In addition, these iPAQ databases will contain the case management information that tracks the events that are recorded as each case is worked. The GSS provides several levels of data QC and built-in QC checks that control operations of a given Field Interviewer. The details of these checks and processes can be found in the field manuals and the *GATS Core Questionnaire Programming Specifications*. This document contains specifications at the column/variable level for each data field in the core instrument.

The following sections describe the resources and procedures that are required to collect the data from the individual iPAQs and merge it into a master database that reflects all the data collected from all the field staff.

4. PSU/Regional Data Management

4.1 PSU Hardware and Software Facilities

Staff, hardware, and software resources will be required to collect and merge data at the PSU/Regional level. PCs and/or laptops are required at each point or level where data is to be consolidated. A modern PC or laptop with 4 GB RAM, a disk drive with at least 150 GB of storage, and the Windows XP operating system are sufficient to process samples of 12,000 households. A system with a multicore CPU will improve the speed of the merge process when there are many files or large files. The PC and/or laptop must be equipped with an SD card port or SD card reader for transferring data to and from the iPAQs. **Appendix D** provides more detail about the hardware and software of the machines required for merging and processing data at each level of data collection. In addition, a table is provided that specifies hardware and software required for staff in different support roles.

This document does not treat the issue of funding this type of equipment. The decisions as to the source and level of funding will be determined in the proposal process.

4.2 PSU GATS Staff

It is recommended that at least one staff member be assigned to each PSU to manage/maintain iPAQs, provide technical support to interviewers, and perform case assignments and data transfer and consolidation operations. Other duties include data backup, security, and QC. Information about data security and backup can be found in **Appendix E**. Staff must have adequate knowledge of PC operations and database management software. PSU staff will require training on iPAQs and software. PSU staff must ensure that data on the PSU PC/laptop are successfully copied to a backup media, such as CDs or external backup devices, daily. **Appendix B** provides more information on suggested staff skill levels and experience that are required for the various roles of staff supporting GATS tasks.

4.3 Case Assignment and Management

Software is provided to assign, track, and manage completed and follow-up sampling cases. This includes the ability to load cases into iPAQs and to track them to their completion. The Case Management System allows cases to be loaded at training, if available at that time, and for cases to be loaded from simple case files stored on SD cards during field operations. In addition, the Case Management System allows cases to be transferred out from one iPAQ and then transferred in to another iPAQ to facilitate moving cases between interviewers (though this is an optional feature which is not standard for GATS). PSU staff will need training in how to build case files to transfer cases between Field Interviewers and assign new cases to Field Interviewers. The GSS case management module provides tracking information at the iPAQ level, and the File Builder program provides summary statistics on cases as they are aggregated. The GSS system also provides a simple reporting feature that will allow reviewing the status of the survey events and frequency distributions of data from the *Individual Questionnaire*. (Information about File Builder and the reporting software can be found in **Appendix F**.) At the national level, reports can be created and tailored to a given country's needs from the event data that

are supplied as part of the GSS using commonly available reporting software (e.g., SAS, SPSS, and ACCESS).[1]

4.4 Data Transfer and Aggregation from iPAQs

Countries will be provided software to transfer and consolidate data from the iPAQs' SDF/SQL databases into a consolidated SDF/SQL database on the PSU PC or laptop hardware. The following procedure should be followed when transferring data from the iPAQ:

1. Each Field Supervisor will be assigned a supply of SD cards to be used to collect export files from each iPAQ on his/her team.

2. Each day the Field Supervisor will collect export files from each iPAQ, using a different SD card for each day of the week. This will be done using the GSS built-in utility Export Data. This utility in the GSS makes a copy of the iPAQ database and labels it with the Field Interviewer ID and date. This database holds the cumulative data on the iPAQ, so each copy of the data contains all data up to the point of the export. The Field Supervisor will insert his/her SD card into a Field Interviewer's iPAQ and run the GSS Export Data command. The Field Supervisor will repeat this process for each Field Interviewer iPAQ on the team using the same SD card. This will build a folder on the SD card that contains each Field Interviewer's up-to-date data. Each week the Field Supervisor will recycle the week's SD cards to collect the next week's data.

3. At a period to be determined in each country, the Field Supervisor will provide the most recent SD card with its collection of export files for his/her team to the PSU/Regional staff.

4. The PSU/Regional staff will insert the SD card into the PC/laptop and copy SDFs to a pre-established staging folder.

5. The PSU/Regional staff will verify that the appropriate SDF databases were copied successfully to the PC/laptop.

6. The PSU/Regional staff will then run the data aggregation software, review the summary reports, and review the merged database for quality and completeness.

7. The PSU/Regional staff will then transmit the merged file either via secure FTP or other methods to the next level of data aggregation (state, province, or country).

8. The national level staff will use GATS-provided software that allows production of national level, master data sets that involve merging all of the key files collected to produce SAS or SPSS ready data sets for analysis.

This process provides for nightly backup of the iPAQ files and for a brief historical backup (up to a week) of data collected on the iPAQ. Because the export data files are cumulative, there is insurance against the loss of any earlier data because each subsequent export collects all prior data. **Appendices E** and **F** provide operating instructions and screenshots of the tools that support data backup and data review.

---

[1]Use of trade names is for identification only and does not imply endorsement by the U.S. Department of Health and Human Services.

4.5 First-Level Quality Control and Validation

Software—Raw Data Viewer—is also provided to perform first-level cross-check and validation tests on the consolidated database. (Information about Raw Data Viewer can be found in **Appendix F**.) It will allow the review of events and completed questionnaires to search for problems or inconsistent data before transferring them to the country database. It will also generate frequencies, by question, for review. Please refer to ***GATS Quality Assurance: Guidelines and Documentation*** for further information on data quality. In addition, the GSS reporting software allows monitoring of the survey progress and data in the *Individual Questionnaire*. **Appendix F** has more information about the reporting option. Three reports are provided that generate the following information:

- result code status (most recent event) of all cases, by Field Interviewer, for the *Household* and *Individual Questionnaires*

- frequency distributions of items from the *Individual Questionnaire* (overall, by Field Interviewer, or by PSU)

- an exceptions report that lists cases having four or more events but are not in a final status

- response rates for the *Household* and *Individual Questionnaires* (overall, and by Field Interviewer)

4.6 Data Transfer to National Data Center

Data transfers into the National Data Center are similar logically to the process at the iPAQ level. At the Country level, the unit of data merged is not the individual iPAQ databases but rather the PSU/Regional databases, which contain the merged data from multiple iPAQs. Each National Data Center will need a file collection mechanism to collect data from the PSU/Regional levels. This could be manual movement via flash drives or electronic file transfers. Also staff at the national level will need to be charged with the tasks of monitoring the inbound flow of these files into the National Data Center.

It is recommended that countries use a secure file transfer protocol to move the data. A tool that will provide this function is the open source program FileZilla. FileZilla is a fast and reliable cross-platform FTP, FTPS, and SFTP client with many useful features and an intuitive interface. It is available in both client and server versions. (See http://filezilla-project.org/ for details.) The FileZilla client, installed on the PSU/Regional laptop, can be used to move the merged PSU/Regional databases up to the Country level. Training will be provided for setting up and using this software tool.

(NOTE: Countries that chose to use the GSS and iPAQ hardware provided by GATS to collect data, but are required to consolidate and aggregate data at the national/Country level into their own database technologies [such as Oracle, IBM/DB2, SA, and/or SPSS], can convert GSS databases into formats that can be imported into these other database products. Users should not have problems exporting data from the SDF/SQL databases into a plain text format that can then be imported into these other database products for easier integration with the national/country database.)

5. Country Data Management

The National Data Center and the implementing agency will be the starting point and the end point for the survey operations. At the start, it will begin the process by designing and building the sample frame and household lists that define the survey. During survey operations, it will be the central point that monitors survey operations and provides the overall QC focus for the survey. At the end, it will be the repository for all data collected during survey operations. These tasks will require human and machine resources. Countries that do not have available existing computing facilities to consolidate and host GATS data from PSUs should plan to acquire a high-end PC(s) with hardware, software, and performance capacity to host, manage, analyze, and generate GATS reports.

5.1 Survey Design Software

The survey design and sample will be developed at the national level by the implementing agency and its partners. This will require staff that can design and draw a national sample and IT staff resources that can convert the sample design into files suitable for use on the GSS devices (the iPAQs). The initial stage for starting a GSS-based survey is the development of the household-level file that drives the case assignment process. Country-level staff will build this file and prepare it for allocation to the individual iPAQs. In countries where a total file of selected households is not available prior to the start of the survey, procedures will have to be developed that will allow the building of interim case files that will allow loading of the iPAQs with households to be surveyed during survey operations.

5.2 Country GATS Staff

At the Country level, IT staff will be required to manage and support the survey activity. It is recommended that one system and database administrator position be filled. Also one additional person will be required to provide technical support and data management support to the PSU/Regional staff. They will assume the following tasks:

- setup of iPAQs and laptop systems

- modification of questionnaires and testing

- setup of multiple language versions of the questionnaires

- setup of file transfer protocols, hardware, and software

- development of a survey monitoring process and monitoring

- development of household-level sample files

- overseeing master data file construction

- routine data management activity

- running of routine statistical reports for QC activities

- support of training activities

These staff will be required for <u>several months prior to the pretest and through the development of the national database</u>. Refer to ***GATS Quality Assurance: Guidelines and Documentation*** for information on additional QC tasks for the in-country IT staff.

Field Interviewers, Supervisors, and IT Staff Training at the PSU/Regional/National Level

Training and Technical Assistance to GATS Countries

Each country will translate and modify the GATS questionnaire to include its supplemental questions. Training on the use and implementation of the survey questionnaire for the iPAQ administration—and the accompanying documentation—is, therefore, necessary for each GATS country (see ***GATS Data Management Training Guide***).

Training

Staff training at the Country level is to be planned and schedule according to the following objectives:

- Objective 1: Provide in-country staff with assistance to program, test, and maintain an iPAQ-based survey, as needed.

- Objective 2: Review the GATS pretest questionnaire, including question and response wording, probing instructions, and pace of questions.

- Objective 3: Help in-country staff use iPAQ hardware and software, as needed. The pretest should "test" all the iPAQ technical aspects inclusive of data collection, aggregation, and transfer.

- Objective 4: Observe pretest implementation in one rural and one urban environment, if feasible.

- Objective 5: Gather information on the main survey implementation plan and the procedures for use in tailoring technical assistance and training support for the main implementation.

- Objective 6: Debrief WHO, CDC, and other GATS staff on strengths, weaknesses, and recommendations on survey pretest preparations and implementation. The IT support technical staff will prepare a site visit report documenting the IT training.

Pretest Site Visit

To ensure the pretest is implemented on schedule, numerous activities must be accomplished prior to the pretest site visit. These activities should be completed 2 to 3 weeks prior to the pretest site visit. Presite visit activities are documented in the ***Presite Visit Checklist.***

Exhibit 5-1 displays the flow of the pretest activities.

Exhibit 5-1.   Pretest Activities
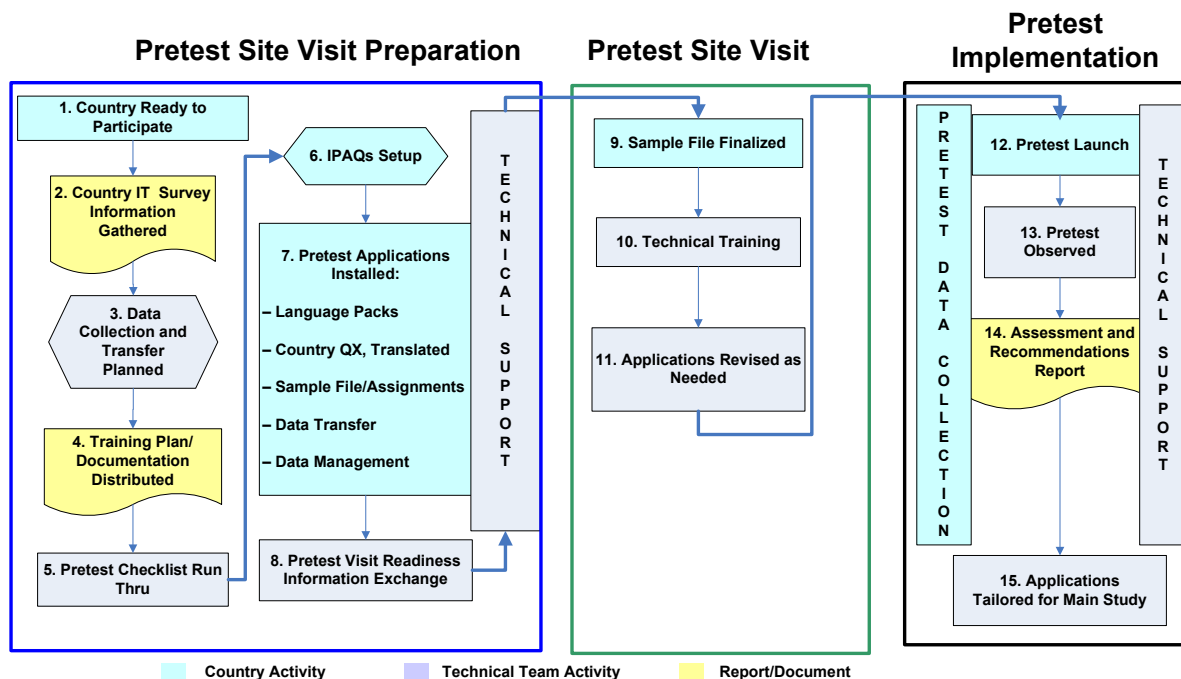
Ongoing technical assistance to the host country survey staff will be provided. Support, to be provided in English, will include telephone and e-mail communications. To facilitate communication, several technical support tools will be used, such as remote access to share screens and a mechanism to exchange files. These support tools will be installed and tested during pretest activities. Technical support activity will be logged and reviewed to search for common problems or emerging issues during survey operations. Regular feedback will be provided to CDC, CDC Foundation, and WHO on technical support activities.

Appendix A: Glossary of Commonly Used Terms and Definitions

Capacity Building: The process of developing and strengthening the skills, abilities, processes, and resources that organizations and communities need to survive, adapt, and thrive.

Data Governance: The process by which an organization manages the quality, consistency, usability, security, and availability of the data.

Data Aggregation: The process of combining data files at various levels (e.g., over Field Interviewers, over Field Supervisor, over regions).

Data Management: Processes for preparing and managing the data inclusive of sample assignments, transfer, and aggregation of the collected data, monitoring and reporting of collected data; based on country's survey operations and IT infrastructure.

Data Transfer: Method of moving data between devices and levels (e.g., from iPAQ to field Supervisor laptop or other data centers).

Data Stewardship: The identification of positions in organizations that are accountable for the definition, production, and usage of the data.

Designer: A GSS questionnaire design program for preparing and/or modifying a survey instrument. The Designer allows the creation, deletion, and modification of questions in two languages in the GSS without writing programming instructions. This application is a PC-based program.

Export: A GSS program used to export data from an IPAQ to an SD card.

File Builder: A GSS data aggregation program that runs both on the iPAQ and on a PC for use in combining Field Interviewer files at each designated field level.

GATS: Global Adult Tobacco Survey.

GSS: General Survey System, a questionnaire development and presentation system engine that allows defining of data collection forms on a standard desktop PC and execution of these data collection forms or questionnaires on the iPAQ.

Guideline: A way of implementing the principle, sometimes with additional information supplied.

iPAQ: Hewlett-Packard handheld device.

Levels of Data: Points of data combination such as Field Supervisor, PSU, region, or National Data Center.

Node: A term referring to a PC—laptop or desktop—machine assigned at a specific level.

SD Card: A Secure Digital card, a type of flash memory card format.

SDF: Secure Digital Format file (SQL Server Compact Edition).

Standard: A general principle that needs to be followed. Standards can be voluntary or mandatory, and as technology and needs change, become superseded.

Surveillance System: Surveillance is the ongoing systematic collection, analysis, and interpretation of outcome-specific data for use in the planning, implementation, and evaluation of public health practice. A surveillance system provides the functional capacity for data collection and analysis and the timely dissemination of information derived from these data to persons who can undertake effective prevention and control activities.

Xmit: A GSS data transmission program that allows bidirectional movement of data, program updates, and control information to and from the iPAQ over dialup, wireless, or wired Ethernet.

Global Adult Tobacco Survey (GATS)   A-2   **Data Management Implementation Plan**
Version 2.0—November 2010   Appendix A: Glossary of Commonly
Used Terms and Definitions

Appendix B: Recommended Staff Position Descriptions

This appendix contains a list of responsibilities, skills and qualifications, role components, and background for each of the proposed data management staff positions. The positions described are:

- IT Specialist/DB Admin and

- Data Manager/Statistical Programmer.

B1.   Information Technology Specialist

Responsibilities for this position(s) encompass supporting the Global Tobacco Surveillance System (GTSS) Global Adult Tobacco Survey (GATS) data collection and management activities and providing technical assistance. In this capacity, the position will perform the following activities:

- Act as the main in-country technical reference for GATS, serving as the primary point of contact for all country iPAQ instrument/questionnaire and iPAQ hardware issues.

- In collaboration with country focal point, act as the liaison with international partners and the implementing agency (inclusive of any contractors) to facilitate all IT activities.

- Coordinate and collaborate with all partners at all levels for the implementation of GATS.

Implementing Agency Support Activities

- Attend GATS IT training as provided by the technical team and the international partners, becoming proficient in the iPAQ usage, software programming, and supporting applications aspects.

- Assist the implementing agency in installation and preparation of programs and procedures for the use, storage, and management of GATS data.

- Assist the implementing agency in identification, procurement, installation, and operation of supporting National Data Center (NDC) electronic equipment such as laptops and supporting software.

- Prepare the iPAQ with the needed software for the implementation of GATS and support all functionality of the use of iPAQs, including data transfer from field staff to a centralized location.

- Provide input for developing ongoing data collection and management activities and effective information dissemination within the country.

- Provide support for data aggregation into the NDC and at other levels, as required.

- Support/perform data dissemination from the implementing agency to the international partners.

- Provide software and hardware technical assistance in the implementing agency for all GATS data collection and management activities.

Field/Operational Support Activities

- Assist in training field staff in using the iPAQ and ensure staff expertise in using it.

- Support the nationals in the technical use of the iPAQs.

- Solve problems that arise in the use of the iPAQs and supporting applications at the national and Country levels.

- Assist in monitoring site adherence to data collection protocol(s), including security and confidentiality.

Programming Activities

- Oversee instrument/questionnaire testing activities prior to release to ensure instrument and data specifications are met.

- Write software programs needed to produce reports and provide support for data export and dissemination as needed.

- Working from the GATS questionnaire specifications, write programming instructions to perform instrument/questionnaire modifications/additions.

Skills

This position demands the following skills:

- strong communication skills—both written and verbal preferred—and the ability to communicate in English

- experience in technical support activities; desire to work in a team-oriented environment

- experience in data collection instrumentation software highly desirable

- experience in Microsoft office products in a Windows XP environment

- experience in SQL server and statistical software (SPSS, SAS, Stata, EPIInfo) preferred

- strong troubleshooting skills

- education or coursework that required the development or adaptation of computer programs and systems and provided knowledge/work experience equivalent a plus

Background

If possible, this position should reside within the implementing agency. Staff will direct the development and operation of the country's GTSS GATS system. Activities include instrumentation and system development, installation, modification, maintenance, monitoring, and/or integration of computer operating systems, applications, and databases to meet GATS requirements and needs.

B.2   Data Manager

The Data Manager will typically be responsible for performing data management tasks such as data validation, cleaning and editing, tracking data collection, monitoring data-related activity including data transfers, maintaining the surveillance data repository, and producing reports and developing and maintaining documentation. This position will ensure the completeness, accuracy, and reliability of surveillance data.

Duties include:

- assisting in the planning and implementation of the data management activities and functions of a large-scale surveillance system;

- working to improve the quality of the data being collected, producing summaries and listings of the data, and testing new processes and components for the management of the system;

- using reports to track progress and ensure that timeliness and quality expectations are met;

- assisting in the standardizing of data management procedures, and reviewing and revising procedural manuals and standard operating procedures as needed;

- responding to data requests when needed, interacting with other staff, and assisting in the preparation of training materials and training activities; and

- working collaboratively with other staff members including country focal point, principal investigator/Survey Director, and survey field staff, statisticians, other Data Managers and IT specialists.

Position Justification

The Data Manager position(s) focuses on routine data management activities specific to maintaining the GATS database(s), reviewing data, reporting, and grantee monitoring. The position is the backbone of the day-to-day data management processing and activities for the system.

Education, Experience, and Skills

- Bachelor's or master's degree in a public health-related science field, computer science, or other information technology specialty preferred with a minimum 6 years' experience in research, specializing in data management, or a least 2 years' experience in clinical data management

- Previous experience as Data Manager

- Intermediate skill level in statistical software such as SAS, SPSS, Stata, and MS Office products, including MS Excel, MS Word, and MS PowerPoint

- Working knowledge of other computer software and software programming, database systems, data collection systems, and data verification and validation techniques

- Excellent verbal and written communication skills

- Good problem-solving skills and the ability to work under pressure and meet deadlines

Appendix C: GSS Setup Activities

Exhibit C-1.   IT Specialist GSS Setup Activities

Step	iPAQ	Laptop
1	Inventory, set up, and test iPAQs	Run installation program to install GSS software and required Microsoft components from provided software distribution
2	If not factory settings, perform factory reset	Check that the following Microsoft products are installed: — MS Active Synch — MS SQL CE — MS SQL native Client — MS.NET Framework
3	Build Field Interviewer list	Check that C:\GATS and C:\Gats_Folders trees contain full GATS GSS software set
4	Label SD cards	Test laptop configuration to see that all GATS Developers' Tools components are available and working: — Designer — Edit DB tables — Case file tools — Data aggregation — SDF generator — Help files
5	Build SD card with GSS files: — App Programs — GenRebuld	
6	Put SD card in iPAQ to run Build Update config file for backups	
7	Make sure iPAQ menu includes: — CMS — Backup SD card	
8	Prepare SD cards with GATS_Casefile.txt for loading cases	

Appendix D: Host Country Required Hardware and Software

The software requirements presented below represent required versions and levels unless indicated that later versions will also be acceptable. The GSS tools have been tested on these specified software versions as indicated below. The hardware requirements are presented as minimum requirements.

D.1   Model A—Full Network

A. Full Network—Web Model

Functionality

Fully networked model provides Web-based:

- case assignment via network upload and download,
- survey monitoring,
- upload of field data, and
- survey monitoring reports.

Requirements

All staff members require access to Internet:

- National Data Center (NDC):

    – *one or more modern Windows XP Service Pack 3 machines (standard laptop <$1,000), 4 GB RAM, 250 GB hard disk*

    – *backup system/devices*

    – *Internet connection*

    – *.NET Framework 3.0 or later*

    – *SQL Server Compact Edition 3.5 or later*

    – *MS Office, with Access version 2003 or later*

    – *memory card reader—reads SD and compact flash cards*

    – *Microsoft Web Server—MS IIS, ASP.net*

    – *Microsoft Active Sync Version 4.5*

    – *secure FTP client and server (sFTP) (e.g., FileZilla)*

    – *encryption software*

    – *expertise to maintain Web site and FTP server*

    – *provide technical support to Regional staff*

- Regional staff and field staff:

  – *one iPAQ per Field Interviewer, with additional iPAQs (1 for every 10 iPAQs in the field, or 10%) to be used for repair or replacement*

  – *access to wireless, wired, or modem link to the Internet*

D.2 Model B—Partial Network

B. Partial Network—Some Field Phone/Internet Connectivity

Functionality

- In-country system hosts consolidated database at NDC.

- Case assignment performed by memory card.

- Interviewers export questionnaire data to memory cards.

- Supervisors or Regional Supervisors collect cards.

  – *consolidate cards to common SQL database*

  – *generate reports*

  – *upload data to NDC*

- NDC combines SQL files across PSUs/Regions.

  – *reporting*

  – *survey monitoring*

Requirements

- NDC and Regional level:

  – *One iPAQ per Field Interviewer, with additional iPAQs (10%) to be used for repair or replacement*

  – *Two SD cards per iPAQ*

  – *Windows XP Service Pack 2 (standard laptop <$1,000), 4GB RAM, 250 GB hard disk*

  – *.NET Framework 3.0 or later*

  – *backup system/devices (USB memory keys or disks)*

  – *FTP server to collect Regional data at NDC; FileZilla recommended*

  – *FTP client on all systems, FileZilla recommended*

  – *SQL Server Compact Edition and Management Studio version 3.5 or later*

  – *MS Office, with Access version 2003 or later*

  – *Internet access*

  – *PC with card reader—reads SD and compact flash cards*

– Visual Studio 2005

– Microsoft Active Sync Version 4.5

– aggregation software

– utilities for viewing data

– software to make flat files

D.3   Model C—Card Based

C. Card-Based

Functionality

- In-country system hosts consolidated database at NDC.

- Case assignment performed by memory card.

- Interviewers export questionnaire data to memory cards.

- NDC combines SQL SDF files from iPAQ-level files:

   – reporting

   – survey monitoring

Requirements

- NDC:

   – one or more Windows XP machine (standard laptop <$1,000), 4 GB RAM, 250 GB hard disk

   – .NET Framework 3.0 or later

   – backup system/devices (USB memory keys or disks)

   – SQL Server Compact Edition version 3.5 or later

   – MS Office, with Access version 2003 or later

   – Internet access

   – PC with card reader—reads SD and compact flash cards

   – flat File Generator

   – Microsoft Active Sync, version 4.5

- Hardware and software required at aggregation sites (PSU Regional offices):

   – Windows XP Service Pack 2 (standard laptop <$1,000), 4 GB RAM, 250 GB hard disk

   – .NET Framework 2.0

   – backup system/devices

   – secure FTP server to upload data; FileZilla recommended

   – SQL Server Compact Edition

       – MS Office, with Access

       – Internet access

       – PC with card reader—reads SD and compact flash cards

       – Microsoft Active Sync Version 4.5

       – aggregation software

       – utilities for viewing data

       – software to make flat files

- Hardware and software required by GSS Developers:

       – Windows XP Service Pack 2 (standard laptop <$1,000), 1 GB RAM, 250 GB hard disk

       – .NET Framework 30 or later

       – backup system/devices

       – SQL Server Compact Edition, version 3.1 or later

       – MS Office, with Access

       – Internet access

       – PC with card reader—reads SD and compact flash cards

       – Visual Studio 2005 with SDK's for Windows Mobile 5

       – Microsoft Active Sync Version 4.5

       – designer software

       – aggregation software

       – utilities for viewing data

       – software to make flat files

- Interviewers/Supervisors/Field Supervisors:

       – Use one iPAQ per Field Interviewer, with additional iPAQs (10%) to be used for repair or replacement.

       – Obtain extra SD cards as needed (cards are shipped to the NDC), with a recommended minimum of two SD cards per handheld.

       – Use mail/courier to send iPAQ SD cards to NDC.

       – If Internet is available at Regional office, data should be sent via secure FTP to NDC.

Global Adult Tobacco Survey (GATS)   D-4   **Data Management Implementation Plan**
Version 2.0—November 2010   Appendix D: Host Country Required
Hardware and Software

Exhibit D-1.   Summary of Required Hardware and Software

Hardware or Software	NDC	Regional	Field Supervisor/ Supervisor	Developer
Laptop PC running Windows XP (see above for details) and MS.NET Framework 2.0	1 or more	Each region	Each Field Supervisor or Supervisor	Each developer
PC with card reader—reads SD and compact flash cards	X	X	X	X
Backup systems (USB hard drives or memory sticks)	X	X	X	X
Internet access	X	X	X (if uploading files)	X
Microsoft Office with Access	X	X	X	X
Secure FTP client*	X	X	X	X
Secure FTP server*	X	Maybe		
Visual Studio 2005 with SDK's for Windows Mobile 5				X
Microsoft Active Sync Version 4.5*	X	X	X	X
SQL Server Compact Edition and Management Studio 2005*	X	X	X	X
Aggregation software	X	X	X	X
Data viewer software	X	X	X	X
Software to make flat files	X			
GSS Designer software				X

*Indicates no cost software either available free from Microsoft or open source.

Appendix E: Security and Backup

Data security and data backup are key activities and requirements for successful data collection. The systems described in this document have built-in security and backup components and support extended procedures where needed. Data security, as defined for this document, is the process by which data on a device or in a file, are protected from access by unauthorized users. Data backup is the process by which data are protected from loss due to hardware or software failures, equipment damage, or equipment loss by making external copies of the original data.

E.1 Security

The systems described in this document support data security by applying several layers of control over GATS data. The GATS databases stored on both the iPAQs and the aggregated databases stored on aggregation devices are encrypted using the built-in Microsoft database encryption that is part of the SQL Compact Edition tool set. Hence data are encrypted while at rest on the laptops, while in intermediate storage on SD memory cards, and while in storage on the iPAQs. The encryption password is known only to system staff supporting the applications and is not entered during normal operations, hence does not need to be known by field workers. In addition, entry into the GATS applications on an iPAQ is protected by a required password. If deemed necessary, optional logon security is available in the iPAQ operating system that requires individual passwords be entered before providing access to the iPAQ. It is expected that laptop systems used for GATS will require usernames and passwords to login to the laptop operating system. The recommended data transmission tools specify an encrypted transmission protocol for electronic movement of the files between aggregation sites.

E.2 Data Backup

The GATS iPAQs have built-in and recommended data backup features. The iPAQ system has built-in data backup using an installed memory card on the iPAQ. Each field worker's iPAQ will have a >=1 GB SD memory card installed. The GSS Case Management System will make a backup of the live, in memory, database every time the field worker exits the Case Management System. In addition, the standard protocol requires that Field Supervisors make backups of the iPAQ data from each Field Interviewer's iPAQ every evening after a day's work (see section on iPAQ data collection). It should be noted that the iPAQ memory system is nonvolatile so that there will be no data loss even if the iPAQ battery loses its charge.

The recommended hardware configuration for devices used as aggregation stations includes several features to support backups. First, the recommended systems are laptop computers, which have built-in battery systems that will provide a minimum degree of battery backup and power management. Second, it is recommended that each laptop have a backup device and a protocol for backup. The recommended devices for backup are USB memory flash drives (with 2 to 4 GB capacity) for systems that will hold smaller amounts of data (Field Supervisor-/PSU-level machines) and USB portable disk drives (150 to 250 GB capacity) for systems that will have larger amounts of data and more complex software installations (Regional-level or National-level machines).

Appendix F: Developers' Tools for Data Monitoring and Aggregation

F.1   Data Viewer Software

View Data—Data Viewer

Overview

The RawDataViewer is a basic tool for inspecting SDF data. It offers the following options:

- View CaseIDs

- View Field Interviewer IDs

- View events data

- View responses data

- View tabulations of the question responses
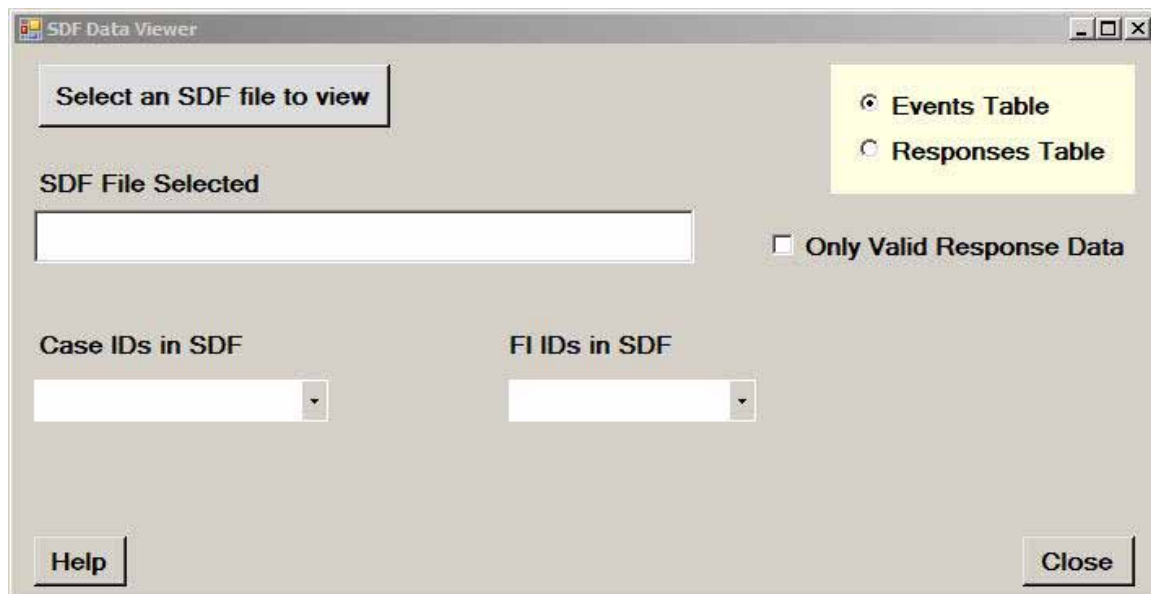
- Export the data to CSV

Terminology

- FI: Field Interviewer

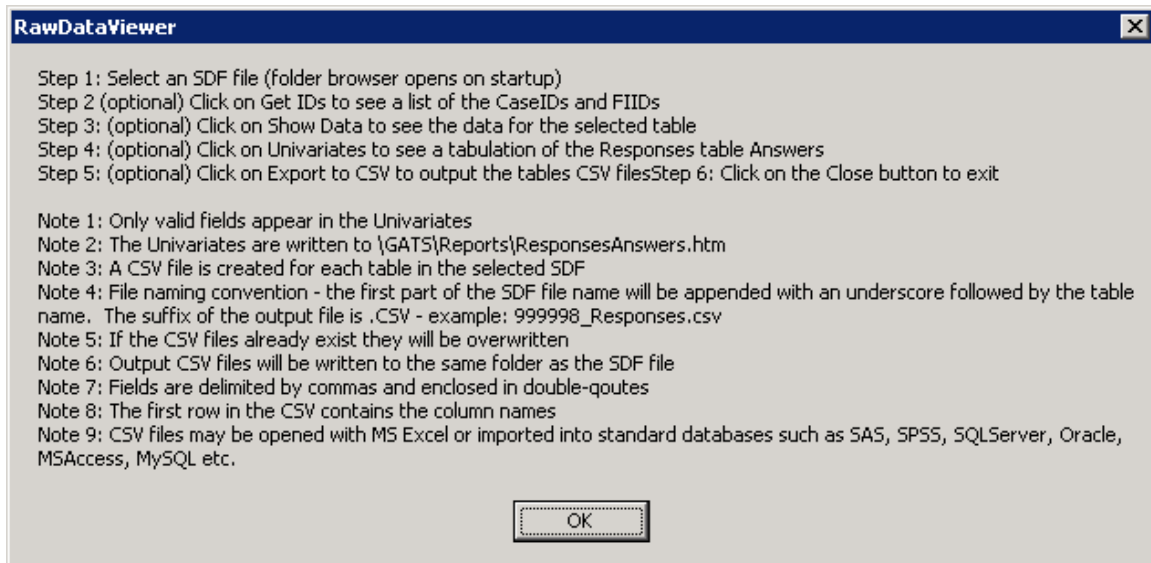- SDF: Microsoft SQL Server Compact (SQL CE) data base file

Process

The RawDataViewer tool is to be used on an ad-hoc basis to inspect data.
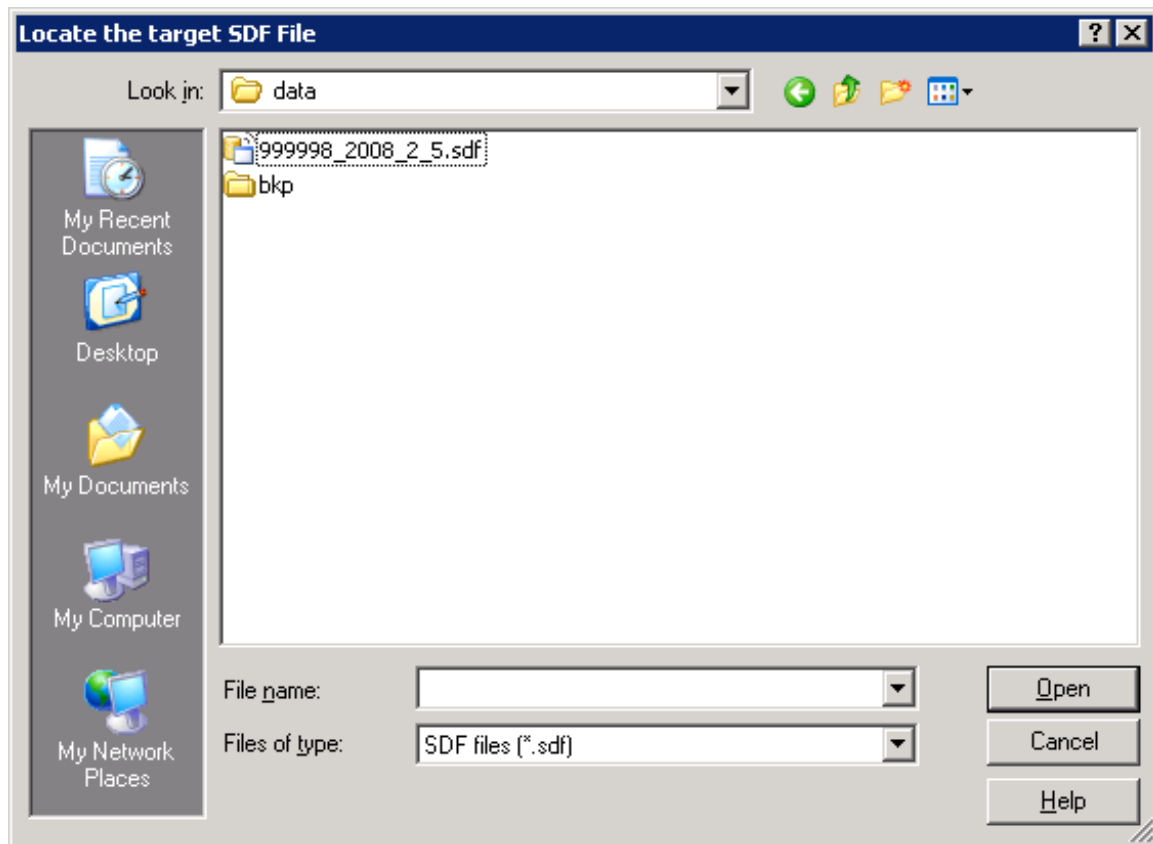
Usage

1. In the GSS Developer Tool set, you invoke the RawDataViwer from the Data Aggregation Menu subitem View Data.

2. Click Help for detailed instructions and notes.

RawDataViewer ✕

Step 1: Select an SDF file (folder browser opens on startup)
Step 2 (optional) Click on Get IDs to see a list of the CaseIDs and FIIDs
Step 3: (optional) Click on Show Data to see the data for the selected table
Step 4: (optional) Click on Univariates to see a tabulation of the Responses table Answers
Step 5: (optional) Click on Export to CSV to output the tables CSV filesStep 6: Click on the Close button to exit

Note 1: Only valid fields appear in the Univariates
Note 2: The Univariates are written to \GATS\Reports\ResponsesAnswers.htm
Note 3: A CSV file is created for each table in the selected SDF
Note 4: File naming convention - the first part of the SDF file name will be appended with an underscore followed by the table name. The suffix of the output file is .CSV - example: 999998_Responses.csv
Note 5: If the CSV files already exist they will be overwritten
Note 6: Output CSV files will be written to the same folder as the SDF file
Note 7: Fields are delimited by commas and enclosed in double-qoutes
Note 8: The first row in the CSV contains the column names
Note 9: CSV files may be opened with MS Excel or imported into standard databases such as SAS, SPSS, SQLServer, Oracle, MSAccess, MySQL etc.

[OK]

3. Click Select a data source directory.

4. Click Open.

Global Adult Tobacco Survey (GATS)   F-3   **Data Management Implementation Plan**
Version 2.0—November 2010   Appendix F: Developers' Tools for
Data Monitoring and Aggregation

5. Click the **Show Data** button.

The show data button displays the data from the responses or events table, as selected. Users can also select a specific Case ID or Field Interviewer ID to filter the data view.

6. Click the **Univariates** button.

This option produces frequency reports for all QIDs that appear in the responses table.

Global Adult Tobacco Survey (GATS)   F-4   **Data Management Implementation Plan**
Version 2.0—November 2010   Appendix F: Developers' Tools for
Data Monitoring and Aggregation

7. Click the **Export to CSV** button.

This option creates CSV files for the following tables in the selected SDF file:

   – *Responses*

   – *Notes*

   – *INI*

   – *DUEVT*

   – *DU*

   – *AggLog*

   – *Addresslog*

   – *These files will be placed in the C';\Gats|Folders\Data_Files folder*

8. Click the **Close** button.

Configuration

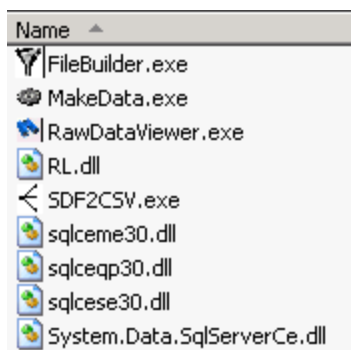Hardware:

- Laptop with 1 Gb RAM and 40 Gb of disk space

File structure:

C:\

Software:

C:\GATS\bin

F.2   Aggregate Data—File Builder Tool

Overview

File Builder is a PC-based program that combines Field Interviewer iPAQ data files. File Builder should be used for aggregating data from the individual iPAQ through as many field levels as are implemented for a given country. The File Builder program facilitates collecting and combining the iPAQ files with the intermediate aggregated files into one cumulative master data file residing on a PC.

File Builder works on SQL Server Compact Edition secure digital format files (SQL CE SDF), which are supported on both the iPAQ and Windows PC platforms. The GSS software has the capability to export SDF files, capturing all the internal data files on the iPAQ. These exported files can then be aggregated together at various data collection points within a country to build a master data file that contains all the questionnaire response data, event data, notes, dwelling unit, and address changes for the survey.

This section covers the use of the File Builder tool from application setup to closing, so that iPAQ data may be aggregated across all levels. Specifically, this section addresses the following:

1. Using File Builder

2. Setting up File Builder on a PC

3. Running File Builder

4. Output generated by File Builder

a. Reports

b. Master Data File

5. Closing File Builder

6. Next steps

1.   Using File Builder

Field Supervisors, Regional Supervisors, State Coordinators/Supervisors, and the National Supervisor will use File Builder to aggregate data from each field level below them. For example, a field supervisor with several field Interviewers reporting to him/her could use File Builder to aggregate the data from each of his/her field Interviewer's SD cards into the master data file on the Field Supervisor's laptop. The Field Supervisor's master data file will then be input to the Regional Supervisor's File Builder process. The Regional Supervisor's master data file will then be input to the State Supervisor's master data file. Finally, the State Supervisor's master data file will be input to the National Data Center's (NDC's) File Builder process. In the end, all data from all sources will be aggregated into the NDC's master data file using File Builder. **Exhibit F-1** illustrates the general process involved in data aggregation at the lowest level.

Exhibit F-1.   Aggregation Diagram
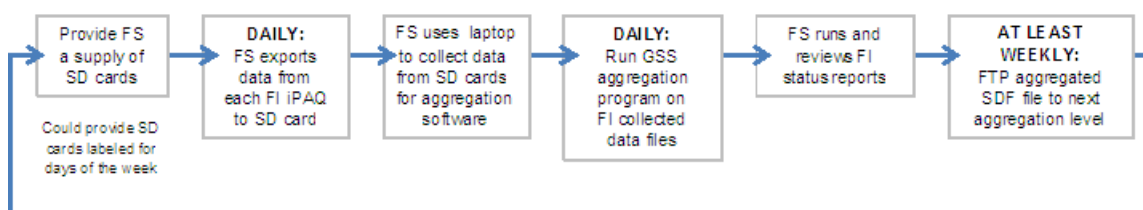
Note: FS is Field Supervisor and FI is Field Interviewer

2.   Setting Up File Builder on a PC

For most countries, Field Supervisors, State Coordinators, and the NDC are expected to use a PC for data aggregation. The File Builder tool is for use on a PC/laptop and requires a one-time installation.

The following PC hardware is the minimum required to install and run File Builder:

- a laptop with 1GB RAM and 40GB of disk space

- an SD card reader

- 1 SD card

The setup steps for the PC are:

    Step 1: Run the supplied installation utility.

    Step 2: Ensure that the SDF data to be aggregated are readily accessible and reside on the PC.

    Step 3: Update a machine identification file, called a node file, on the PC.

Copying File Builder to the Target PC

For use at the PC level, you must copy the File Builder software to the C:\GATS\ folder so that the folder structure, files, and executable code exist. The software is supplied by an install utility. The IT/Computer Specialist who sets up the PCs for field use should use the utility to install the software. It will result in the following set of folders:

Global Adult Tobacco Survey (GATS)   F-7   **Data Management Implementation Plan**
Version 2.0—November 2010   Appendix F: Developers' Tools for
Data Monitoring and Aggregation

SDF Field Interviewer Files to be Aggregated

The data files to be aggregated must exist as SDF files in an accessible location on the PC running the File Builder tool. The input sdf files (from the iPAQs) must be in a different PC folder than the folder that contains the aggregated master file. Data must be exported from the iPAQs to the PC ahead of the File Builder data process execution.

The iPAQ SDF data file transfer (or export) can be performed as described in the following steps:

1. Field Interviewer to next level (e.g., Field Supervisor) aggregation:

   – *Field Supervisor removes the operational SD card from the Field Interviewer's iPAQ.*

   – *Field Supervisor inserts the aggregation SD card into the Field Interviewer's iPAQ.*

   – *Field Supervisor runs Export to copy data from the iPAQ to the aggregation SD card.*

   – *Field Supervisor removes the aggregation SD card from the Field Interviewer's iPAQ.*

   – *Field Supervisor re-inserts the Field Interviewer's operational SD card into the Field Interviewer's iPAQ.*

   – *Field Supervisor inserts the aggregation SD card into the laptop SD card reader.*

   – *Field Supervisor runs File Builder.*

2. All other aggregation

   – *Assumption: SDF data have been transferred from lower levels to the current node (FTP, diskette, CD, SD card, memory stick, etc.) to a specific folder on the node (e.g., C:\GATS\Incoming).*

   – *Node administrator runs File Builder.*

For detailed instructions on using the iPAQ GSS Export utility, see the **GATS Field Interviewer Manual**, **Section 9C.1.**
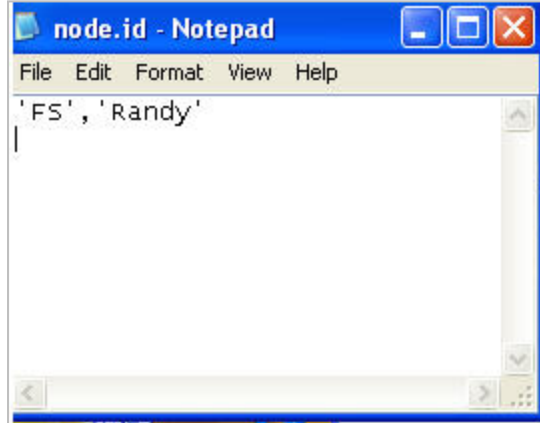
Creating a Node File

A node file is a necessary piece of information you set that identifies the aggregation level/person/machine and is key to the aggregation building process. In the C:\GATS folder, a file called node.id must be edited using a Windows editor such as Notepad, Textpad, or any other text-file editor. The file is to contain the information used to index the aggregation log entries. The node.id file must contain a unique name for each node id, or the data will not be moved up to the next aggregation level. Once data have been aggregated, the node.id file should not be modified.
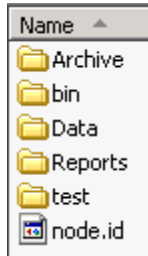
The node.id file must contain a single record with the information starting in column 1, row 1, specifying two pieces of information: the node level designation, such as the Field Supervisor, and the identifier/name of the node.

Single quotes and comma are required for all fields entered (e.g., 'Field Supervisor', 'Randy').

Global Adult Tobacco Survey (GATS)   F-8   **Data Management Implementation Plan**
Version 2.0—November 2010   Appendix F: Developers' Tools for
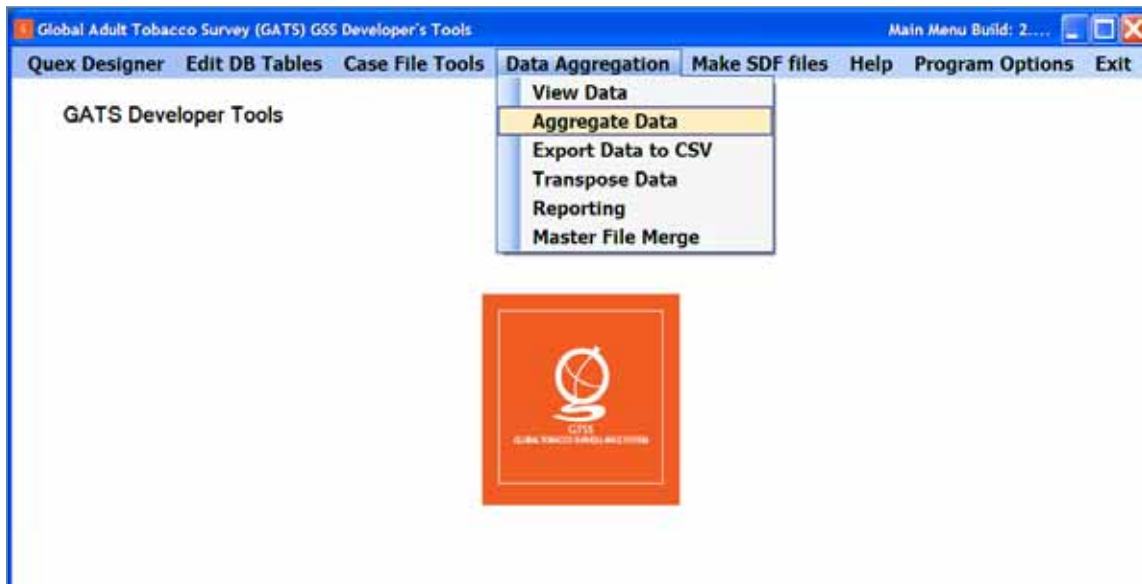Data Monitoring and Aggregation

Example of a node file:

The Node File should be named **node.id** and be saved to the GATS folder:

3.   Running File Builder

To initiate the File Builder you may invoke it from the GSS IDE Main Menu data aggregation menu item—Aggregate Data or double-click the **FileBuilder.exe** file in the C:\GATS\bin folder.

Note the following points about File Builder:

- File Builder combines the SDF files to be aggregated in **ascending order by filename**.

- File Builder aggregates five tables from the input SDF files; these are:

  – *AddressLog (address change data)*

  – *DU (Case information, address, sample info, etc.)*

  – *DUEVT (event data)*

  – *Notes (interview notes about the cases)*

  – *Responses (questionnaire data)*

- File Builder creates a record-level "key" for the records in the input tables using the CaseID.

- FileBuilder does not allow for duplicate records in the aggregated file, based on the CaseID.

- File Builder creates a single, aggregated master SDF file and does not alter the input SDF input files. Ideally, the input folder should contain the most recent files that need to be aggregated. A separate folder may be created that contains the SDF files that have already been aggregated, so they are not aggregated again. The user should note that exported SDF files from an iPAQ are always a complete copy of the data so that information that was exported in earlier files will also be present.

- The master SDF file always resides in the GATS\Data folder.

- Every step in the aggregation process is cumulative.

At this point in the process, the following screen is displayed, indicating that the user should enter a single data source directory in which all the data files to be aggregated reside.
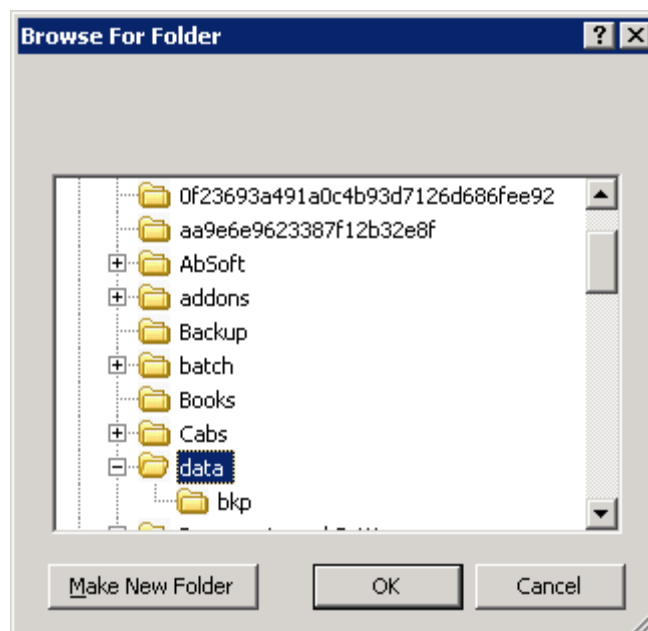
File Builder Help

A Help screen is available within File Builder. This screen provides some instructions and notes. To access the Help screen, click the **Help** button, and the following screen appears.
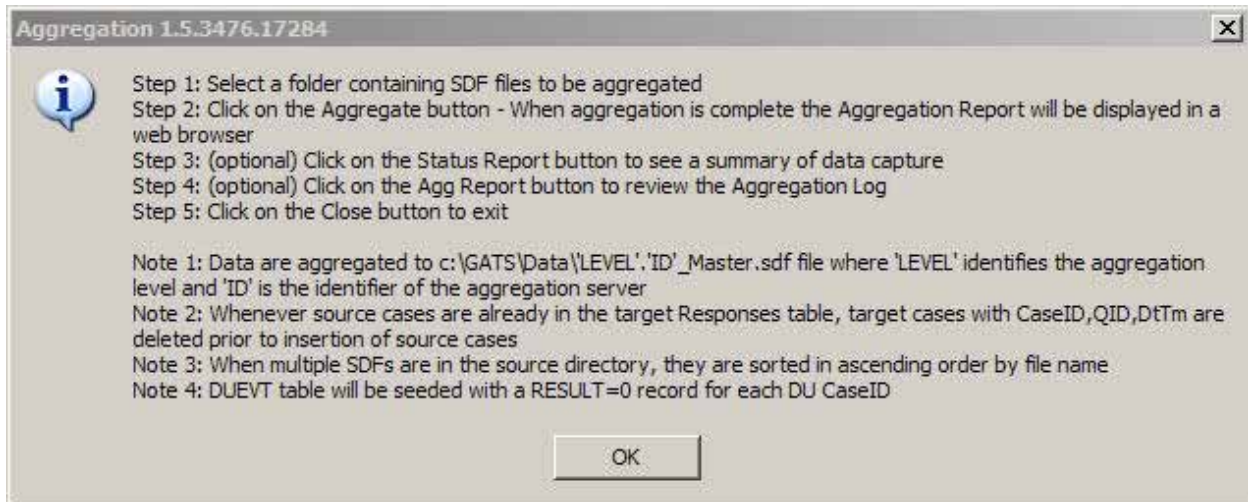
Click on **Select a data source directory,** and browse for the directory/folder in which the iPAQ data were previously saved. Although it is not necessary to maintain the iPAQ data files in a certain folder, it is advisable to store them in the C:\GATS\Data folder on the PC.

Once the directory containing the SDF data has been identified, click on the folder name, and then click **OK**.

The SDF files residing in the PC folder that you selected are then listed on the File Builder screen.

Click the **Aggregate** button. Doing so will aggregate/combine all the files in the folder. A screen is displayed indicating the tables that have been read in from the aggregated SDF files.

Global Adult Tobacco Survey (GATS)
Version 2.0—November 2010

F-12

Data Management Implementation Plan
Appendix F: Developers' Tools for
Data Monitoring and Aggregation

4. Output Generated by File Builder

a. *Reports*

Once the data have been aggregated successfully, File Builder can provide several reports, detailed below.

Aggregation Summary Report—directly after the aggregation summary screen is displayed and the aggregation is complete, a summary table showing the log of the aggregation is generated. The contents of the generated file **aggrpt.htm** are displayed using the default Windows HTML editor on the PC. See **Exhibit F-2** for a screenshot of the Aggrpt file.

Exhibit F-2.   Aggregation Summary Report (file Aggrpt.htm)

GATS Data Aggregation Summary

Date: 4/27/2008 3:37:30 PM

AggDateTime	AggTarUnit	AggTarUnitID	AggSrcUnitID	AggTable	AggRowsTotal	AggRowsInserted
4/27/2008 3:37:00 PM	FS	Randy	999990	DU	24	24
4/27/2008 3:37:00 PM	FS	Randy	999990	Notes	2	2
4/27/2008 3:37:01 PM	FS	Randy	999990	DUEVT	3	3
4/27/2008 3:37:01 PM	FS	Randy	999990	AddressLog	0	0
4/27/2008 3:37:01 PM	FS	Randy	999990	Responses	41	41
4/27/2008 3:37:12 PM	FS	Randy	999998	DU	24	0
4/27/2008 3:37:13 PM	FS	Randy	999998	Notes	2	0
4/27/2008 3:37:13 PM	FS	Randy	999998	DUEVT	3	0
4/27/2008 3:37:13 PM	FS	Randy	999998	AddressLog	0	0
4/27/2008 3:37:14 PM	FS	Randy	999998	Responses	41	0
4/27/2008 3:37:26 PM	FS	Randy	999998	DU	24	0
4/27/2008 3:37:27 PM	FS	Randy	999998	Notes	0	0
4/27/2008 3:37:27 PM	FS	Randy	999998	DUEVT	7	7
4/27/2008 3:37:27 PM	FS	Randy	999998	AddressLog	0	0
4/27/2008 3:37:30 PM	FS	Randy	999998	Responses	598	598

Design Source Quick View

The aggregation summary report can be generated by clicking the **Agg Report** button that is displayed after files have been successfully aggregated.

Status Report—by clicking the **Status Report** button on the File Builder aggregation screen, a status report summarizing the data from the SDF files will appear (file statrpt.htm). See **Exhibit F-3** for a screenshot of the statrpt.htm report.

Exhibit F-3.   Status Report (Statrpt.htm)

GATS Household Screening Status

Date: 6/30/2010 15:00:36

Unit	ID	Cases	Unworked	Pending	Final	CompletedInts	Date of Most Recent Data
FI	999998	14	10	0	4	4	6/30/2010 14:05:38
Totals		14	10	0	4	4	
%		100	71.4	0	28.6	28.6	

- Note: Final counts include completed interviews, unworked is event code 0, pending is event codes >0 and <200, CompletedInts is event code 200; counts in the Cases column is sum of unworked, Pending and Final.

GATS Individual Questionnaire Status

Date: 6/30/2010 15:00:36

Unit	ID	Cases	Unworked	Pending	Final	CompletedInts	Date of Most Recent Data
FI	999998	4	0	0	4	3	6/30/2010 14:22:49
Totals		4	0	0	4	3	
%		100	0	0	100	75	

- Note: Final counts include completed interviews, unworked is event code 0, pending is event codes >0 and <400, CompletedInts is event code 400; counts in the Cases column is sum of unworked, Pending and Final.

The status report can be generated by clicking the **Status Report** button that is displayed after files have been successfully aggregated.

This Data Aggregation Summary Report provides information about the results of the aggregation. Each row shows the results for an aggregation unit (typically a Field Interviewer) and a specific database table. The table is denoted in the AggTable column, the number of rows in the input table is shown in the AggRowsTotal column, and the number of rows inserted from that table is shown in the last column AggRowsInserted. The report above shows in the first line that for FI 999990 and the DU table there were 24 rows and all 24 rows were inserted into the master aggregation file. In cases where there are fewer rows inserted than were in the inbound table, the rows already exist in the master table and were not updated.

The Household Screening Status report is a summary of the status of all the cases in the aggregated data file. For each aggregation unit (typically a Field Interviewer) the total number of cases is reported for that unit in the Cases column. Then the total is broken out into the various status categories. The status columns (Completed Ints, Unworked, Pending, and Final) depend on the most recent result code for the case. The result codes are defined in the INI table of the CMSDB data base. The typical GATS values are shown in **Exhibit F-4**.

Global Adult Tobacco Survey (GATS)
Version 2.0—November 2010

F-15

Data Management Implementation Plan
Appendix F: Developers' Tools for
Data Monitoring and Aggregation

Exhibit F-4.   Typical GATS Values

INI		
VariableName	VariableValue	Comment
EvtcodeComplete	200,400	Result codes that are complete quex
EvtCodeFinal	200,201,202,203,204,205,206,208,209,999, 400,401,402,403,404,407,408,409	Result codes that are final status codes
EvtCodePending	0,102,103,104,105,106,108,109,302,303, 304,307,308,309,887	Result codes that are pending status codes
EvtCodeScreened	200	Result code indicating screening complete and one person selected for more data collection

Unworked cases have a most recent result code of 0. Note that the final status includes the cases that are completed interviews. The Cases column is the total for a given aggregation unit and should equal the sum of the Pending, Unworked, and Final columns. In the status report for the *Individual Questionnaire*, cases are filtered. Only households that have a result code 200 from household screening (screening interview complete and eligible HH member selected) are shown in this table.

b.   File Builder Output Data File—The Master Aggregated SDF

Once the aggregation has successfully completed, File Builder will create a single combined SDF data file, considered the master data aggregation file, located on the PC in the \GATS\Data folder. The first time File Builder is run, the master SDF file is named using the **node.id** information (e.g., FS_MyID_Master.sdf). The original SDF file(s) used in the aggregation will reside as it is in its previously saved location.

Browse the C:\GATS\Data folder to find the master SDF file that the aggregation process has created.
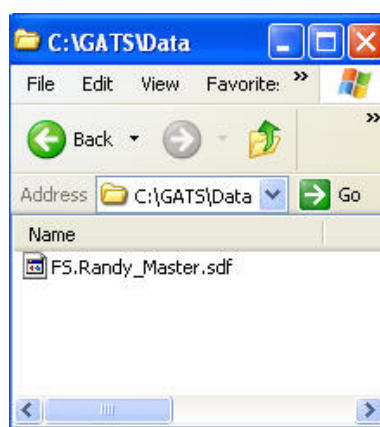
5.   Closing File Builder

If you click the **Close** button at any time during the aggregation process, the File Builder application closes, and you will exit from the application.

Global Adult Tobacco Survey (GATS)
Version 2.0—November 2010

F-16

Data Management Implementation Plan
Appendix F: Developers' Tools for
Data Monitoring and Aggregation

Once the aggregation process has run to a successful completion and the aggregation and status reports have been generated, click the **Close** button to exit File Builder.

6.   Next Steps

The next steps to be performed depend on the current level of aggregation. The master SDF file created by the aggregation process may now be sent up to the next level (via FTP or other means) and aggregated there, as needed.

F.3   Transpose Data Utility

This utility provides the capability of taking the aggregated Responses data and creating a variable list and files for Survey 0 and Survey 1. It is a VB.net program that reads the MDB for a given survey and makes a variable list from this MDB. Using the variable list as a template, it extracts the response data from the aggregated SDF file and makes a single data record for each unique ID in the SDF file.

The opening screen for the transpose utility is shown in the following screenshot. You select the survey 0 or 1 by editing MDB path block. Six data blocks on the screen allow users to enter variables that drive the construction of the variable list and flat file. The values shown in the screenshot are the suggested defaults. The definitions of the variables are as follows:

- Loop limit—maximum number of "instances" of a variable in a loop for the GATS *Household Questionnaire* (20 is used)

- Text Max width—maximum width, in characters, of a text type item

- List Max Width—maximum width, in characters, of a list type item

- Num—maximum width of numeric fields

- # Dec—number of decimal places
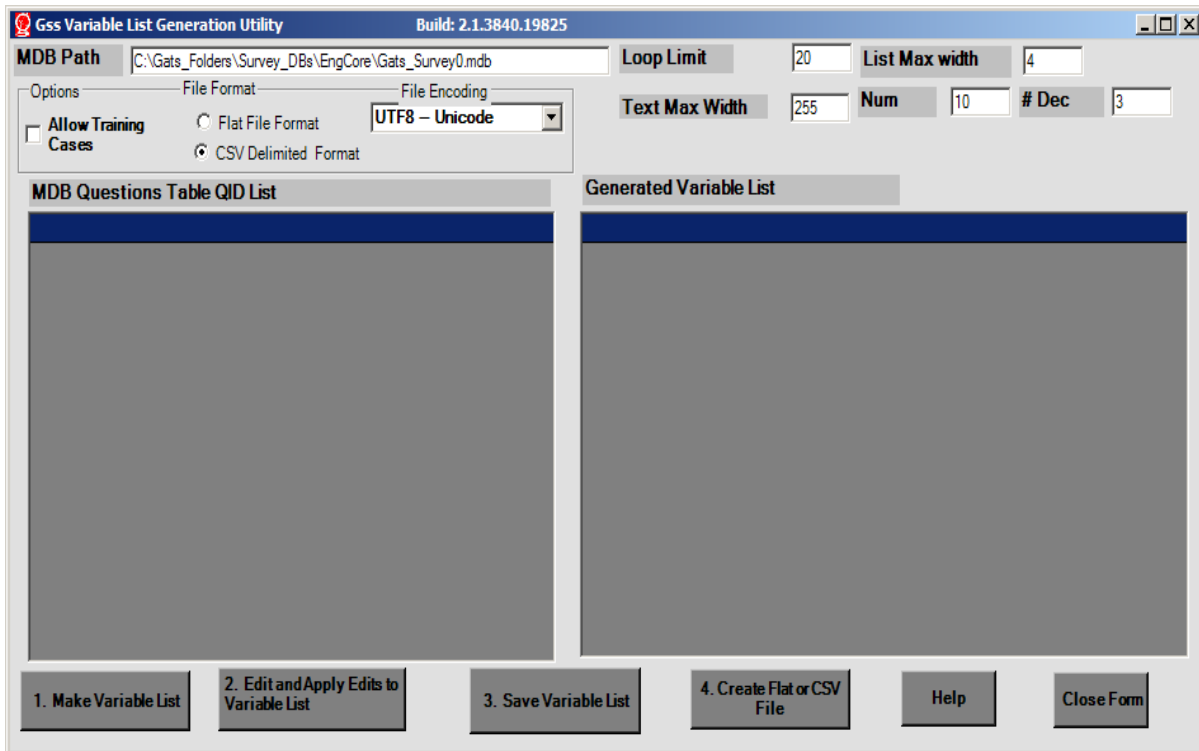
You can select whether to allow training cases by checking the option box. You can also select the format for the output file (flat file or CSV file) and its encoding. Again the selected options in the figure are the recommended defaults.

To start the process, tap the **Make variable** list button. The following screen displays:

Global Adult Tobacco Survey (GATS)
Version 2.0—November 2010

F-18

Data Management Implementation Plan
Appendix F: Developers' Tools for
Data Monitoring and Aggregation
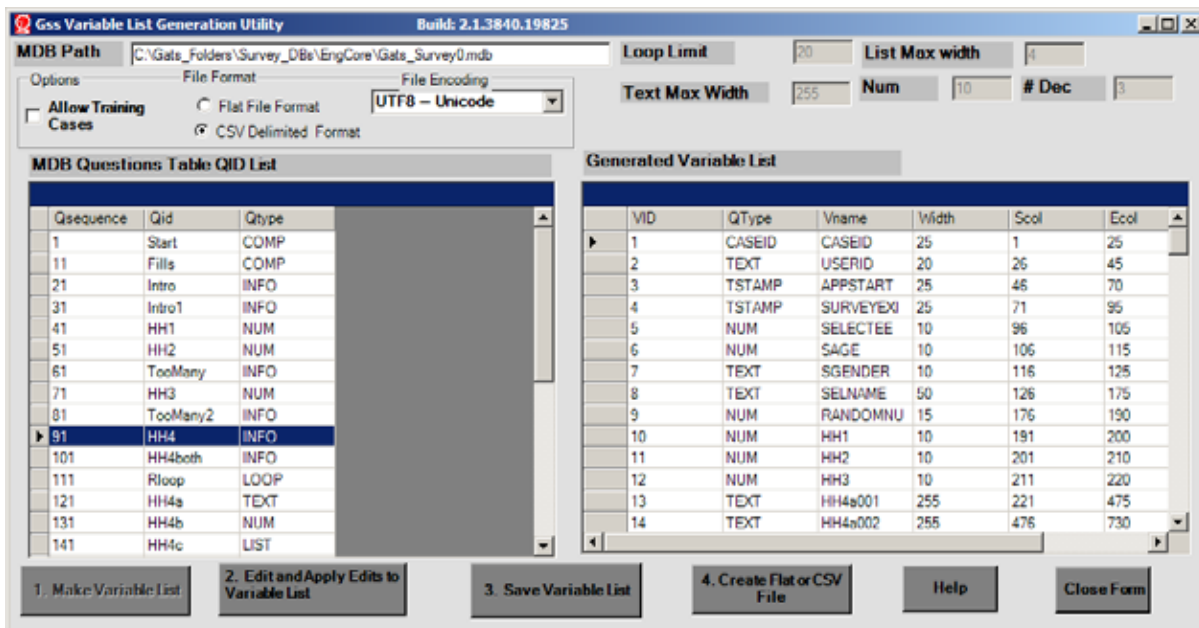
You can edit the Generated variable list (the grid on the right of the screen) to change the widths of item on a one-by-one basis if desired. If you edit the list, you can tap **Edit and Apply Edits to Variable List** (button #2), and the edits will be applied. If you wish to save the variable list in a file, tap **Save Variable List** (button #3) and the utility prompts you for a storage location for the file. It saves the list as a comma-delimited file with six variables: sequence number, variable type, variable name, width, start column, and end column:

Sample of variable list file

```
1,CASEID,CASEID,25,1,25

2,TEXT,USERID,20,26,45

3,TSTAMP,APPSTART,25,46,70

4,TSTAMP,SURVEYEXIT,25,71,95

5,NUM,SELECTEE,19,96,105

.

.

.

.
```

Finally, to create the flat or CSV file, tap **Create Flat File or CSV File** (button #4). This action creates a file using the variable list derived earlier. It makes one record for each unique ID in the responses table for the selected survey, 0 or 1. The method of making the transpose is as follows:

- Step 1: First select the distinct IDs in the aggregated file.

- Step 2: Then for each ID found in Step 1

- Step 3: Select the **valid data and sort** it by QID, date time, prikey (an auto-number sequence in the responses table).

- Step 4: For each QID variable found in Step 3

  – *Place each variable in the designated record spaces as dictated by the variable list, and overwrite older data with the latest data if there are multiple instances of data for the same QID.*

  – *If a CSV file is desired, convert the fields in the flat record to CSV format for output.*

The actual SQL select for this operation is as follows:

- "Select * from responses where CaseID='" & ThisId & "' and valid=1 order by qid, dttm,prikey"

Global Adult Tobacco Survey (GATS)
Version 2.0—November 2010

F-19

Data Management Implementation Plan
Appendix F: Developers' Tools for
Data Monitoring and Aggregation

where ThisId is a specific Case ID. After you tap button #4, the system prompts you for a place to store the new file and where to find the input SDF file that contains the data you are transposing.

The Help file for the transpose utility is repeated below.

F.4   Help File for Variable List and Flat or CSV File Generation

This utility is designed to reformat the data in the GSS Responses table (the answers to questionnaire questions); in individual SDF files or aggregated SDF files, into a form that is easier for analysis software to process. The GSS responses table stores one row for every answer to a question and it contains backed over data as well as current data. This format is the complete data record and audit trail but is hard to work with in standard data analysis packages like SAS or SPSS. This utility allows users to convert the responses data into a format that is easier to work with in SAS or SPSS. It transposes the question by question row oriented format of the responses table into a distinct flat record for each case. That is, it makes a single record, for each case, by taking the **most recent, valid response** to each question. If a question was not asked the field for it is filled with blanks. The utility is driven by the mdb file for the GSS questionnaire specifications for either Survey0 or Survey1. That is, the utility uses the GSS mdb questionnaire specifications to create a list of all possible questionnaire variables and variable types. The screenshot below, **Exhibit F-5**, shows the startup screen. The Help button on this screen will invoke this help file.

The user can select the format for the output file (flat file or delimited (CSV)) and the data encoding (various forms of Unicode (UTF7, UTF8, UTF16 little or big Endian, UTF32) or ANSI). The flat file format is one single, character record, with fixed width fields defined, according to the variable list in the first step of the program. The delimited format is comma-separated fields, one field for each row in the variable list. Text fields with embedded commas have leading and trailing "(double quote) characters added to them. Text fields that have embedded "(double quote) characters in them have the " removed. The default format is CSV and the default encoding is UTF8. Unicode (UTF8) and CSV-delimited formats will work in SPSS, and either format will work in SAS. Unicode is required for character sets that require 16 bit characters (e.g., Russian and Chinese), and we have tested the UTF8 format in SAS and SPSS.

The user can specify the limits and maximums for loops, text question field widths, list question field widths, and width and number of decimals for numeric questions. Also the user can allow training case data or not depending upon the state of the check box Allow Training Cases, if checked training cases (IDs that start with XX) will be processed, otherwise, they will be ignored.

Exhibit F-5.   GSS Variable List Utility
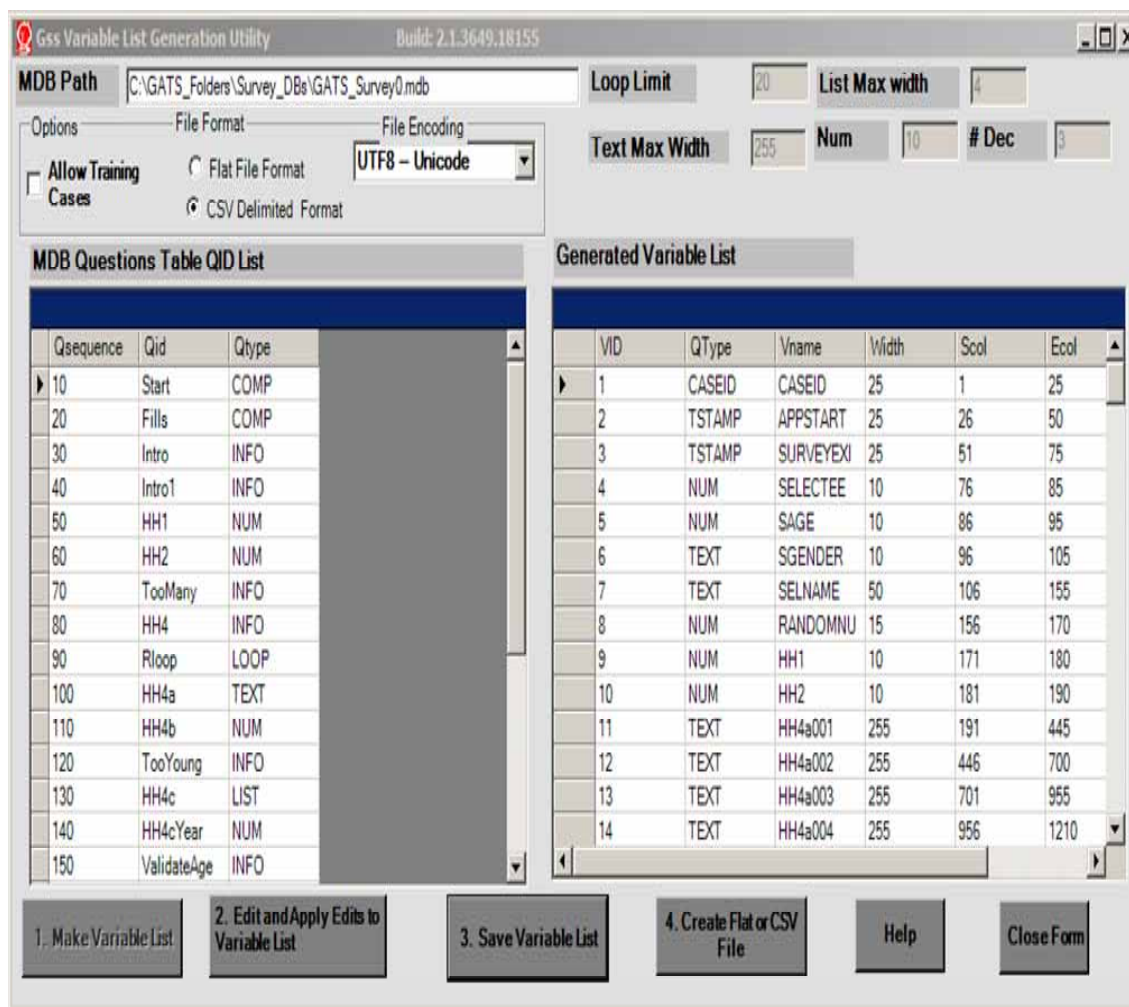
This software will generate two files for the user. It will generate a variable list from the input MDB file that is referenced in the MDB Path box. This is done by tapping the **Make Variable List** button. This variable list can be edited to change widths of the columns. You can save the variable list to a file to use for reference or input to statistical software macros by using the **Save Variable List** button. After tapping this button, you will be asked for a file location to store the variable list as shown in **Exhibit F-6**.

Exhibit F-6.   Request for Path for Storing Variable List (.txt) File
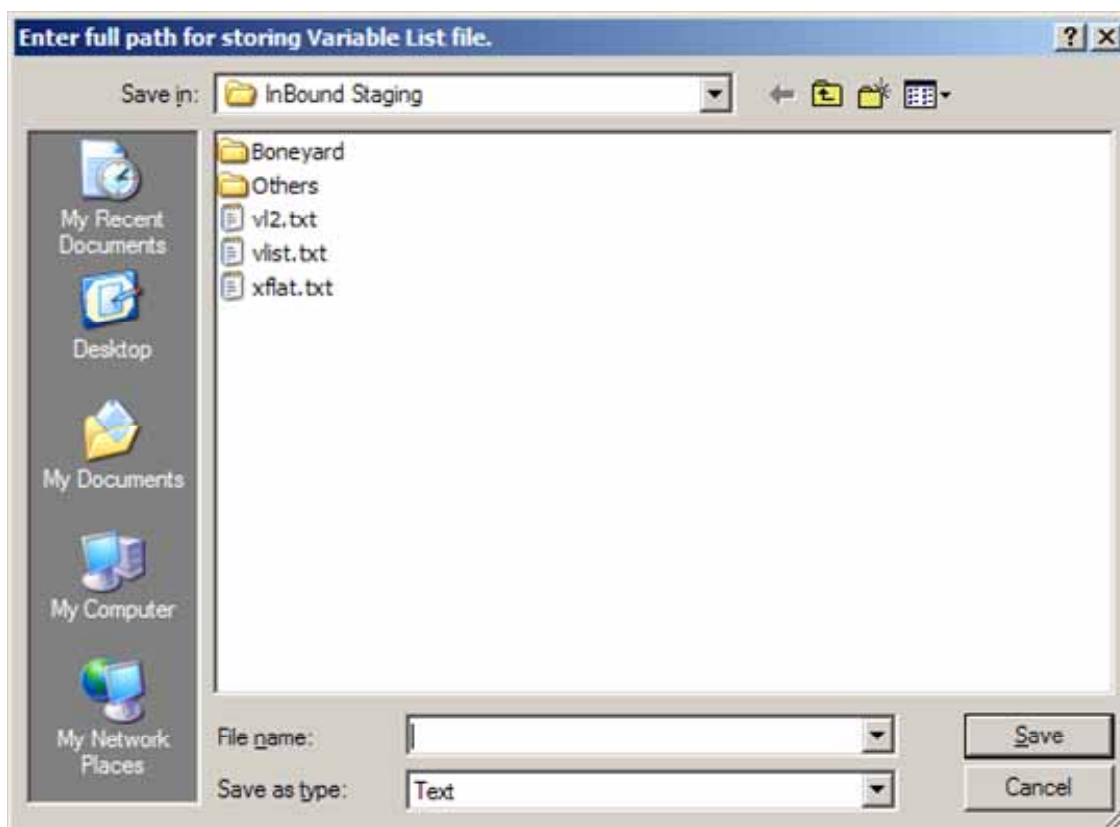
Once edits are complete, you can ask the tool to create a transposed file from a selected SDF file using the variable list that you have just generated as a guide or driver for the file creation. You do this by tapping the **Create Flat** or **CSV File** button. After tapping this button you will be asked to browse to file locations for an output file (**Exhibit F-7**) and an input file that contains the raw data (**Exhibit F-8**).

Exhibit F-7.   Request for Path for Storing Output Data File (flat file or delimited file)

Exhibit F-8.   Request for Path to Input SDF File That Has Raw Data
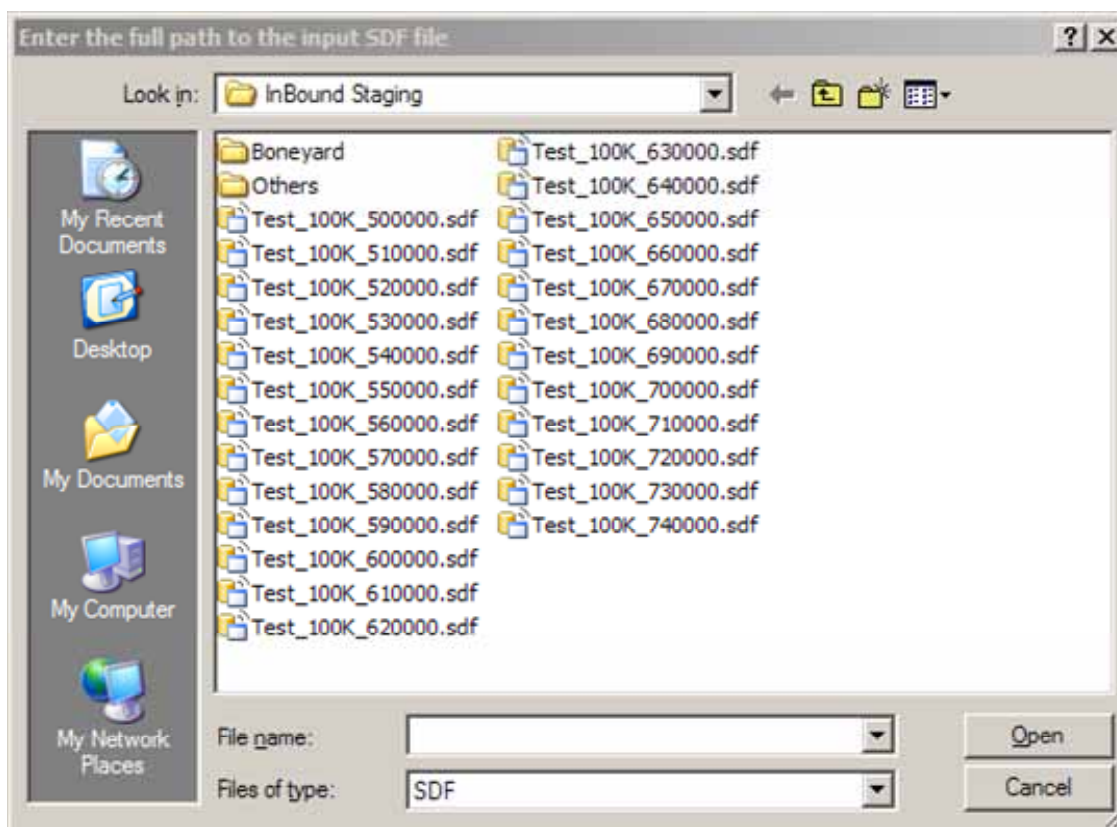
While the transpose is in progress, it will display a case count (updated every 300 cases). Once the output file has been created, it will display this final summary screen. The screen below reports that 193 cases have been created and the last case ID is 100273-01. The user should expect a file of 10,000 cases to take 10 to 15 minutes to complete on a typical laptop. **Exhibit F-9** shows a sample of this progress summary screen.

Exhibit F-9.   Transpose Summary Screen
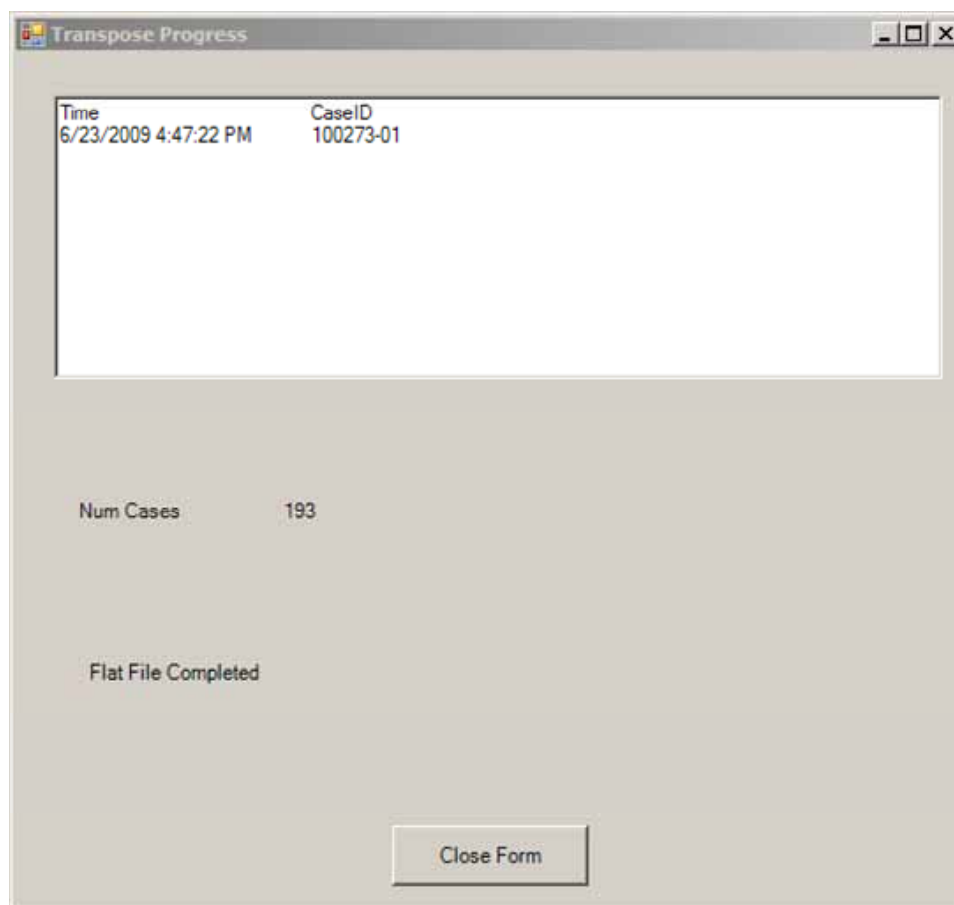
F.5   Using SAS or SPSS

The default file generated by this utility will be a Unicode (UTF8) CSV file. The file will contain one row for each case that was in the input SDF files for the utility. Each row will contain a field separated by commas for each variable in the variable list. If the question was not asked the field will be null (,,). These are standard Microsoft style CSV files and can be easily imported into Excel or Access as well as being used by SAS or SPSS.

F.6   Reporting Utility

The Data Aggregation menu contains a reporting tool that allows users to monitor the content of an aggregated data file for status monitoring and QC purposes. The Reporting menu option brings up a stand-alone utility. This utility allows the GATS manager to create three different status or data reports to view on screen or export to a CSV or text file. The user should point the utility to the Master Aggregated SDF file that he has made by aggregating the individual SDF files from the Field Interviewer handhelds. In addition to the Master Aggregated SDF file, the PC that executes these programs will need current versions of the MDB files that define the surveys on the hard disk and will need to identify the PSU variable in the case DU table. The mdb files will be used to produce labels and formats for the reports, and the PSU variable can be used as a "by" variable for frequencies. The three reports that are available are:

- **Case Status Report**—A Survey Event Status Report based on the latest event or Record of Call code (from the DU Table) that details that status of each Field Interviewer's work load (see **Exhibit F-10** for an example).

- **IQ Frequencies Report**—The second report option allows GATS Data Managers to request descriptive frequency distributions of the data from completed *Individual Questionnaires*. This report will allow managers to use a report filter that controls the level of the report from drop-down lists of PSUs and or Field Interviewers. Also drop-down lists will allow the Data Manager to select up to 20 of the *Individual Questionnaire* variables for a given report and select one of the existing survey languages to use for the labels of question text and response stub text for the report (see **Exhibit F-11** for screenshot of the filter screen and **Exhibit F-12** for an example of the frequency report).

Exhibit F-10.   Filter Screen

Global Adult Tobacco Survey (GATS)   F-26   **Data Management Implementation Plan**
Version 2.0—November 2010   Appendix F: Developers' Tools for
Data Monitoring and Aggregation

Exhibit F-11.   Survey Event Status Report HQ Case Status

Exhibit F-12.   Questionnaire Frequency Distributions

- **Exceptions Report**—The third report lists Field Interviewers who have exceeded a specified number of call records for a case but have not coded the case as final (see **Exhibit F-13** for a sample). This report is based on the result codes in the aggregated DUEVT table. If a Field Interviewer has a case that has four or more record of calls records but is still not in a final status, then that case will be listed in the report. This report will detail separately cases for the *Household Questionnaire* (HH) events and the *Individual Questionnaire* (IQ) events. If cases have HH events but no IQ events, the IQ columns will be blank.

Exhibit F-13.   Field Interviewer Exceptions Cases with No Final Codes and ≥ Four Result codes


These tools should be used by the Data Manager(s) to review and evaluate the survey status and data as the survey operations proceed. The manager(s) should be reviewing the aggregated SDF data files at the level of aggregation for which they are responsible (region or state or nation).

F.7   Master File Merge

The Master File Merge option combines study data into a single comma-delimited file and creates SAS and SPSS input programs capable of reading the newly merged file. To use the Master File Merge, select **Master File Merge** from the **Data Aggregation** menu item.
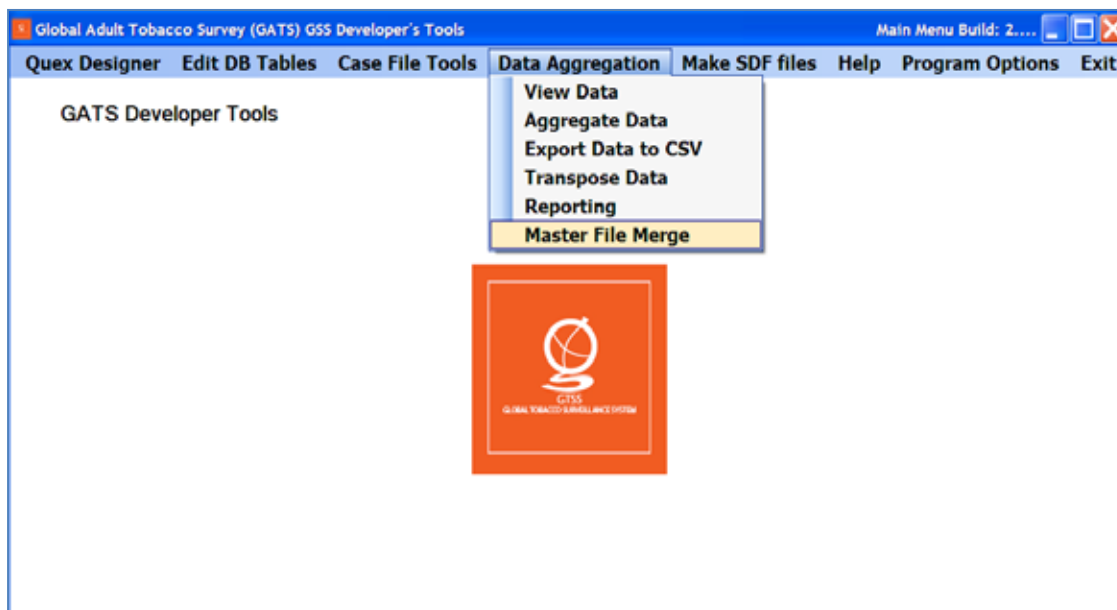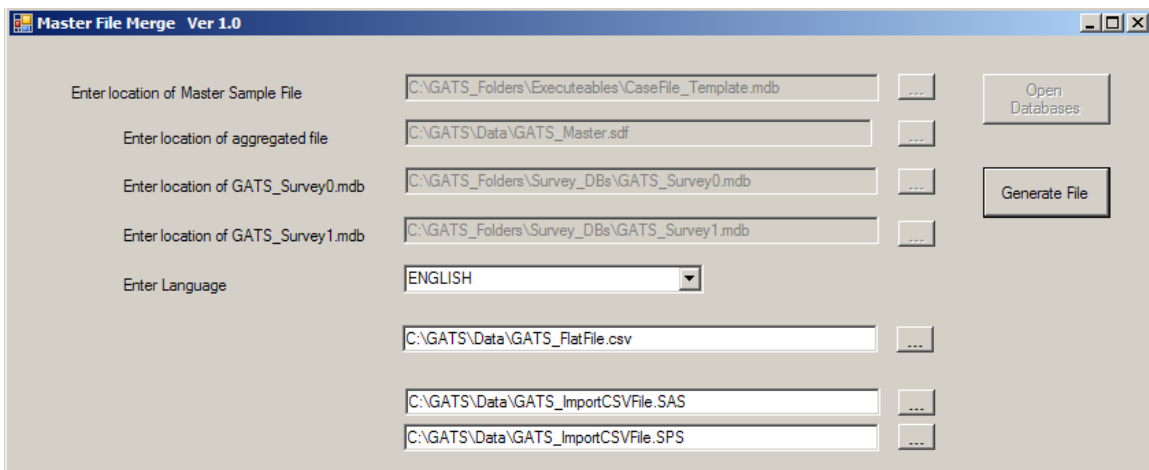
The system has default values for the location of the required files. The files used in the Master File Merge include the Master Sample File, the aggregated SDF file, and the two questionnaire databases (GATS_Survey0.mdb and GATS_Survey1.mdb). The Master Sample File is the MS Access database used to create the case file for the handheld. Additional fields may be added to this database, but the field names, types, and sizes should not be updated. The field description of fields used by your site should be updated so that the descriptive labels produced by SAS and SPSS are correct.
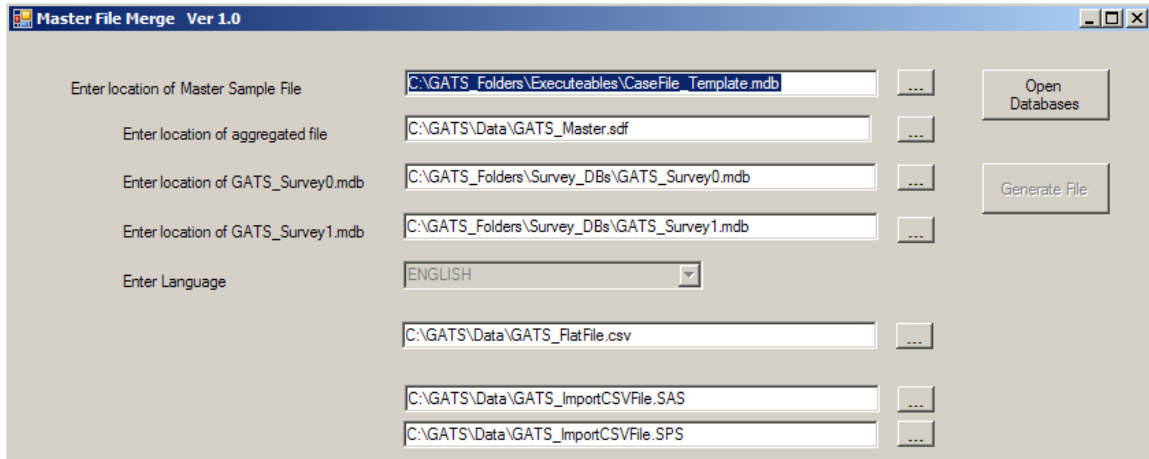

Start by updating the file locations of the master case file, the aggregated file, and the survey questionnaire databases. To change, update the file location or name in the text box or select the "…" button and the system will allow you to select an existing file. Once the files locations and names are correct, select the **Open Databases** button. This will check the file locations and ensure that the system can open the databases. It also gathers information about languages that are available.

After the **Open Databases** button has been pressed and the databases are opened, the option to change or open other databases is not available. The user can now select the language for SAS and SPSS formatted data to display. At any time, the user can select the folder and name of the CSV file, the SAS, and the SPSS input programs that are created.

Global Adult Tobacco Survey (GATS)
Version 2.0—November 2010

F-29

Data Management Implementation Plan
Appendix F: Developers' Tools for
Data Monitoring and Aggregation

Click on the **Generate File** button to create the CSV, SAS, and SPSS files. The CSV will be a comma-delimited file in Unicode (UTF-8) format. This program may take several minutes to run. The hourglass mouse pointer will display while the program is running. Once it is completed, the following message is displayed.
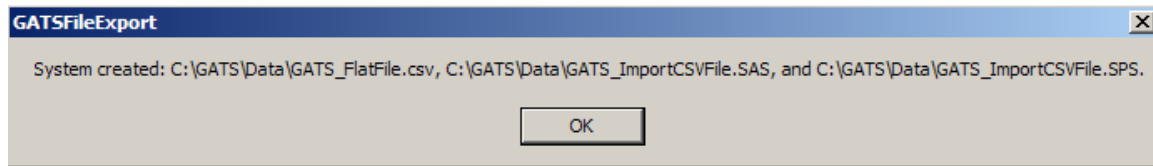
To view the CSV file, a Unicode-enabled text editor such as WordPad is required. The SAS and SPSS files are also in Unicode to ensure that labels and formats are displayed correctly. The SAS and SPSS programs may need to be updated prior to running them to ensure that country-specific options are set.

GLOBAL TOBACCO SURVEILLANCE SYSTEM (GTSS)