# The Arizona Border Study

***An Extension of the***
***Arizona National Human Exposure Assessment Survey (NHEXAS)Study***
***Sponsored by the Environmental Health Workgroup of the Border XXI Program***

# Quality Systems and Implementation Plan for Human Exposure Assessment

The University of Arizona
Tucson, Arizona 85721

Cooperative Agreement CR 824719

**Standard Operating Procedure** **SOP-UA-D-44.0**

**Title:** Operation Manual of the Mass Data Massage Program

**Source:** The University of Arizona

U.S. Environmental Protection Agency
Office of Research and Development
Human Exposure & Atmospheric Sciences Division
Exposure & Dose Research Branch

# REVIEW RECORD FOR STANDARD OPERATING PROCEDURES

| ID NO: UA-D-44.0 | REV: 0 | ISSUE DATE: June 1997 | REVIEW DATE: 05 Aug 97 |
|---|---|---|---|

**TITLE:** Operation Manual of the Mass Data Massage Program

Working SOP ☐
Full SOP ☒

**ISSUED BY:** Health and Respiratory Sciences
The University of Arizona

☒ Original Issue ☐ Revision

| VERIFICATION SCOPE: (check as applicable) ** FIRST REVIEW ** | SAT | UNSAT | N/A | If UNSAT, explain below or on reverse side |
|---|---|---|---|---|
| SPECIFICALLY | | | | |
| Purpose | X | | | |
| Applicability | X | | | |
| Definitions | X | | | |
| References | X | | | |
| Discussion | X | | | |
| Responsibilities | X | | | |
| Equipment | X | | | |
| Sequence of Operation | X- | -X | | see next page |
| Records | X | | | |
| GENERALLY | | | | |
| Technical Approach | X | | | |
| Communication Effectiveness | X | | | |
| QA/QC Requirements | X | | | |
| Other: Appendices J and K | X- | -X | | see next page |

**COMMENTS:** This is an extremely well-written SOP and the flow diagrams in the appendices are excellent. Please see next page, however, for some questions that need to be resolved. Please also see text for some editorial corrections.

***Preparer: Please respond to comments above and/or on next pages.***

☐ **APPROVED**

_____ (Signature) _____ (Date)

☒ **APPROVED WITH REVISIONS ABOVE**

Susan Bishop (Signature) 05 August 1997 (Date)

☐ **NOT APPROVED**

_____ (Signature) _____ (Date)

1095

REVIEW RECORD FOR STANDARD OPERATING PROCEDURES

| ID NO: UA-D-44.0 | REV: 0 | ISSUE DATE: Jun 97 | Operation Manual of the Mass Data Massage Program |
|---|---|---|---|

| FINDINGS/COMMENTS | RESPONSE |
|---|---|
| ¶7.1 (p.8) When does this occur in the sequence of events? Is the program already open on computer? If so, which computer, or does it not make any difference? Maybe 'shell file' is what is unclear. Has this been defined in a previous SOP? Or is it a term in normal computer language?<br><br>pp.14 ff: Does the notation < > indicate press 'Ctrl + letter'? Is this standard computer notation? Or does one just press the 'letter'? Perhaps the reviewer is missing something. Please clarify.<br><br>Appendix J: This diagram appears to have an unending loop. How does one exit or go to 'DONE'? Please explain.<br><br>Appendix K: Encryption of codes to retain confidentiality is extremely important. But the reviewer did not find any discussion in this SOP of how to change the codes. Is this explained in another SOP? Who is allowed to change the codes? The Project Data Manager as per ¶5.2? After a change, who is permitted to know the code? How are they informed? Is each new code saved somewhere in case the Data Manager is gone? Please respond and/or elaborate. | |

| Title: Operation Manual of the Mass Data Massage Program | |
|---|---|
| Document No. UA-D-44.0 | APPROVALS |
| ☒ Full SOP ☐ Working SOP #pages 21 | On Site Principal Investigator: |
| Issue Date: June 1997 | Project QA Director: |
| Revision No. 44.0 | Independent Reviewer: |

| | |
|---|---|
| Revision No: | On Site PI: |
| Revision Date: | Project QA Director: |
| Revision Made: | Independent Reviewer: |
| Revision No: | On Site PI: |
| Revision Date: | Project QA Director: |
| Revision Made: | Independent Reviewer: |

| Distributed To: | Revision No. | | | | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |
| | | | | | | |

Form TP-2

# Operation Manual of the Mass Data Massage Program

## 1.0 Purpose and Applicability

The purpose of this SOP is to describe the operation of the data processing program. These methods will be used for every execution of the data processing program of the NHEXAS project, the Border Study, and other Health and Environment projects.

## 2.0 Definitions

2.1 BORDER STUDY = An alias for "Total Human Exposure Arizona: A Comparison of the Border Communities and the State" conducted in Arizona by the University of Arizona/Battelle/Illinois Institute of Technology consortium.

2.2 DATA = Classified under this word are the following definitions: DATA, ELECTRONIC; DATA, ENTERED; DATA, LOGIC CHECKED; DATA, MISSING; DATA, PHYSICAL; DATA, RANGE CHECKED; DATA, RAW; DATA, VERIFIED; DATA, APPENDAGE, DATA CLEANING, DATA CORRECTION, DATA PROCESSING BATCH, DATA CLEANING BATCH, DATA PROCESSING ERROR, DATA RECORD, DATA VALIDATION.

2.2.1 DATA, ELECTRONIC = Data stored on some type of magnetic or optical (e.g.,: floppy disk, hard disk).

2.2.2 DATA, ENTERED = Electronic data that have been entered for the first time into a computer database. This is the product of "data entry".

2.2.3 DATA, LOGIC CHECKED = Data records that have been checked for and cleared of all apparent logical errors. This is a product of "logic checking".

2.2.4 DATA, MISSING = A datum or data that were applicable to a sample or question, but were not recorded on the physical data form at the time of initial observation; or one or more data records that were applicable to a data processing batch, but were not appended to working or master database.

2.2.5 DATA, PHYSICAL = A datum or data written on a physical data form.

2.2.6 DATA, RANGE CHECKED = Data records where each variable in the record was compared to a pre-established valid range, and adjusted if necessary. This is the product of "range checking".

2.2.7 DATA, RAW = Electronic data that were entered and verified, but not yet cleaned. (See DATA, ENTERED and DATA, VERIFIED).

2.2.8 DATA, VERIFIED = Electronic data that were re-entered into the same table and database into which they were was originally entered, and compared to the original entered values. This is a product of "data verification".

2.2.9 DATA APPENDAGE = The process of adding newly cleaned data cleaning batches to a master database. This is distinct from "data correction". (See DATA CORRECTION).

2.2.10 DATA CLEANING = This is the process of locating and correcting data processing errors (see DATA PROCESSING ERROR below). They can be individual errors in the electronic and physical data, or they can be system level errors in the data collection, packaging, coding, entry, and cleaning procedures themselves. This process is also referred to as "data validation".

2.2.11 DATA CLEANING BATCH = A collection of electronic data, along with their corresponding physical forms. Data cleaning batches are formed after one or more data processing batches (see DATA PROCESSING BATCH below) are scanned. The data cleaning batches are then cleaned (see DATA CLEANING), quality assured (as described in SOP# UA-D-26.X), and appended to the master database (as described in UA-D-27.X). Each data cleaning batch is assigned a numeric descriptor of the form MMDDYY, where MM is the month the batch was created, DD is the day the batch was created, and YY is the year the batch was created. If more than one batch is created on the same day, each batch after the first is assigned a descriptor of the form MMDDYY_N, where N denotes the batch as being the Nth batch created that day.

2.2.12 DATA CORRECTION = A change made to a datum or data within a working or master database. The change can be the modification of the value in one or more data fields, the deletion of one or more data records, or the addition of one or more data records that were missing (see MISSING DATA). This is distinct from "data appendage" (see DATA APPENDAGE above).

2.2.13 DATA PROCESSING BATCH (DP BATCH) = A collection of household packets or physical data forms reviewed for quality assurance and ready for data entry. Each DP batch receives a unique numeric or alphanumeric code that is written on all forms in the DP batch.

2.2.14 DATA PROCESSING ERROR = This is an error that can occur at any level of data processing. It is a procedural mistake, such as a duplicate data record, a typographical error, a logical error, or missing information.

2.2.15 DATA RECORD = In the context of this SOP, this is the set of all electronic data in a database associated with a particular physical form.

2.2.16 DATA VALIDATION = See DATA CLEANING.

2.3 DATABASE = Classified under this word are the following: DATABASE, EMPTY; DATABASE, MASTER; DATABASE, WORKING.

2.3.1 DATABASE, EMPTY = A database structure that contains no records.

2.3.2 DATABASE, MASTER = This is the accumulative database generated from validated data cleaning batches. Newly cleaned batches are appended to the master database.

2.3.3 DATABASE, WORKING = A database earmarked for or in the process of cleaning or quality assurance that contains one or more data cleaning batches. When cleaned, this will be appended to the corresponding master database.

2.4 DUMMY BATCH = A test batch of data with known problems and solutions used specifically for validation/verification purposes

2.5 ERROR = Program control type, informs user of serious problem that requires intervention or correction before work can continue. Subject to validation.

2.4 HEALTH AND ENVIRONMENT PROJECTS (or H&E) = An umbrella title for all projects funded to M.D. Lebowitz and/or M.K. O'Rourke (or their designees) which examine purported or real relationships among environmental factors and any aspect of human health.

2.5 HINT = Program control type, explains the use of a field or control. Supplementary to use of application. Not in specs. Not subject to validation.

2.6 KEY VARIABLE(S) = A variable or set of variables in a data record whose value or combined values make a data record unique from other data records in the same database.

2.7 MASS DATA MASSAGE PROCESS (or MDM) = The data processing program used by NHEXAS Arizona, the Border Study, and other Health and Environment projects.

2.8 NHEXAS Arizona: Acronym for **N**ational **H**uman **EX**posure **A**ssessment **S**urvey, a research project conducted in Arizona by the University of Arizona/Battelle/Illinois Institute of Technology Consortium.

2.9 PROMPT = Program control type, display that asks for user input of a value or serves as a label to a field. Sometimes subject to validation in places where proceeding to operation would not allow for backtracking to pre-operation status.

2.10 TAR = Tape Archive facility

2.11 TARFILE = A file created using *tar*

## 3.0 References

3.1 "The AWK Programming Language." Alfred V. Aho, Brian W. Kernighan, Peter J. Weinberger. ATI Bell Labs/Addison-Wesley Publishing, Co. 1988.

3.2 "The Icon Programming Language." Ralph E. Griswold, Madge T. Griswold. Prentice-Hall, Inc. Co. 1990.

3.3 "TeleForm Standard : User Guide Version 5" Cardiff Software. Co. 1991-1996

3.4 "Sun OS Reference Manual." Sun Microsystems. P/N 800-1751-10. Rev A, 5/9/88

3.5 "SPSS." SPSS, Inc. 1990. Associated Documentation Includes:
(a) "SPSS Advanced Statistics User's Guide."
(b) "SPSS Base System User's Guide."
(c) "SPSS For UNIX: Operations Guide."
(d) "SPSS Reference Guide."

## 4.0 Discussion

Scanning and other technologies facilitate data collection and entry, but there is still a need to ensure data correctness. The *MDM* is an automated, programmatic approach to generating cleaned and quality assured (QA'd) master databases in a time-effective manner. This process includes creating data cleaning batches, which are then cleaned, QA'd, and then finally appended to the master database.

Each type of datafile has an assigned download area. This is where scanned data is saved, and all download data is stored. All new data of a particular type is appended to the download file in this assigned area. At any time, the Data Technicians may choose to create a cleaning batch from the data in this download file. When this is done, the data is formatted to a flat ASCII format and moved to an assigned in-process area, and the original download file is archived.

Each in-process area is dedicated to a specific datafile type, and in each in-process area there is a dictionary corresponding to that specific datafile type. The Data

Technicians will select a particular batch in the in-process area, and the *MDM* will apply the dictionary to that batch, reporting any errors found. So long as errors are found in this way, the Data Technicians will review these errors and suggest corrections. These corrections are recorded in a Change Log, which is itself a runnable script that may be applied to its datafile to obtain the cleanest version of the data. Once the *MDM* finds a data batch to have no errors, e-mail is sent to the Project Data Coordinator informing of the batch being clean. The Project Data Coordinator will then review the changes made. If any changes are rejected, mail is sent to the Data Technicians about the rejected changes. The *MDM* will then apply the dictionary to the data batch. If any errors are found, the batch is returned for further cleaning. Otherwise, mail is sent to the Data Technicians informing that the batch is ready to be QA'd.

When a Data Technician selects a batch to be QA'd, the *MDM* will randomly select 10% of the records. The Data Technician will then go through each field and confirm its validity and correctness. Any changes recommended are then recorded in that batch's Change Log. Once the fields have all been reviewed, the *MDM* sends e-mail to the Project Data Coordinator regarding the QA'd batch. At this point, the Project Data Coordinator may review the new changes as given in the Change Log. A report of any rejected changes is sent to the Data Technicians. If there were fewer than 5% of the fields with errors, the *MDM* sends mail to the Project Data Manager informing of a batch ready to be appended to the master. Otherwise, the 10% QA is considered to have failed and the above procedure is repeated with 50% of the records randomly selected. If this check fails, all files related to the cleaning batch are erased and the data is re-entered.

When the Project Data Manager is ready to append the data to the master database, the cleaning batch, Change Log, and any QA files are archived. The Change Log is then applied to the cleaning batch, and the master database is updated with the cleanest version of the data batch.

To secure the validity of the MDM, a dummy batch is added to every tenth data processing batch and run through each step of the data cleaning process. This is done to ensure that the MDM is functioning as expected.

The MDM encompasses the following procedures:

| Function | Protocol # |
|---|---|
| Correcting Electronic Data | UA-D-25.X |
| Data QA Check Protocol | UA-D-26.X |
| Addition of Cleaned Scannable Data Batches to Master Databases | UA-D-27.X |

The Appendices describe, in a broad scope, the algorithms used within *MDM*. The Appendices are:

| | |
|---|---|
| Appendix A: | Batch Creation |
| Appendix B: | Cleaning of Batches |
| Appendix C: | QA'ing of Batch |
| Appendix D: | Inspecting of a Cleaned Batch |
| Appendix E: | Inspecting of a QA'd Batch |
| Appendix F: | Appending of Data |
| Appendix G: | File Integrity Check |
| Appendix H: | Univariate Check |
| Appendix I: | Bivariate Check |
| Appendix J: | Performing a Global Change |
| Appendix K: | Changing of Encryption Keys |

## 5.0 Responsibilities

### 5.1 Project Data Coordinator

#### 5.1.1 Supervision of Data Corrections

The Project Data Coordinator supervises corrections made to databases by Student Data Technicians. He/she does this by providing solutions regarding unusual or ambiguous data, sometimes in consultation with the principle investigator, the Project Data Analyst, the Project Data Manager, and/or the Project Field Coordinator.

#### 5.1.2 Approval of Data Corrections

The Project Data Coordinator gives formal approval for each correction made to a database by a Student Data Assistant.

### 5.2 Project Data Manager

The Project Data Manager is primarily responsible for master databases. He/she (1) appends QA'd cleaning batches to the appropriate master databases, (2) approves suggested corrections to the master databases, (3) makes any approved corrections to the master databases, (4) makes any necessary global changes or re-codes to the master databases, and (5) makes corrections to the master databases that were generated from the statistical analysis run on one or more derivatives of the master databases. The Project Data Manager may also perform the duties of the Project Data Coordinator and Student Data Technicians as described in this SOP.

### 5.3 Student Data Technicians

5.3.1 The Student Data Technician is responsible to the Project Data Coordinator and/or Project Data Manager.

5.3.2 The Student Data Technician has custody of the data processing batches and data cleaning batches earmarked for or in the process of being cleaned, until they are submitted to the Project Data Coordinator for approval.

5.3.3 The Student Data Technician is responsible for documenting any corrections to the physical data.

5.3.4 The Student Data Technician is responsible for cleaning and QA'ing the cleaning batches.

5.4 Project Data Analyst

The Project Data Analyst is responsible for the analysis of the project data. This analysis includes:

5.4.1 The Bivariate Check: The review of master databases to ensure the existence of forms, which may involve cross-checking master databases

5.4.2 The Univariate Check: The review of outliers in the master database

Any discrepancies and suggested solutions are reported to the Project Data Manager, who will further review the problem and make the appropriate change (if any).

5.5 Project Field Coordinator

The Project Field Coordinator is responsible for providing information regarding the standard operating procedures for the preparation of the household packets, data processing batches, and the collection of field data, when it is needed for the data correction purposes.

5.6 Student Field Technician

The Student Field Technician is responsible for (1) providing specific information about the household packets on which he/she has participated and (2) providing specific information about field data that he/she has helped to collect, when it is needed for data correction purposes.

## 6.0 Materials & Reagents

6.1 H&E Local Area Network
6.2 Data Processing Batches
6.3 Purple Pen

## 7.0 Procedure

### 7.1 Preparations

Add the following line to users' shell file allowing for access to program.
Line is : alias data '/rsc53/NHEXAZdata/cleaning/src/data'

### 7.2 Steps Followed

A. Data Technicians
   1. Log into UNIX environment.
   2. Type "data" to run the MDM.
   3. Select form type:
      a. If a field form is desired, the user will enter "1". The *MDM* will then prompt the user for a specific form. The user will then enter:

         A. "1" for Household Summary Sheet
         B. "2" for PM Sampling
         C. "3" for PID Sampling
         D. "4" for Soil Sampling
         E. "5" for Floor Dust Sampling
         F. "6" for Surface Sampling
         G. "7" for Sentinel Sampling
         H. "8" for Personal Air Sampling
         I. "9" for Sample Tracking
            If this option is selected, the user will be prompted to select either the header or the detail groups. The user will enter "1" for the header, and "2" for the detail group.
         J. "0" to return to the previous menu

      b. If lab data is to be cleaned, the user will enter "2". The *MDM* will then prompt the user for a specific form. The user will then enter:

         A. "1" for 25mm and 37mm pre-weighing
            If this option is selected, the user will be prompted to select either the header or the detail groups. The user will enter "1" for the header, and "2" for the detail group.
         B. "2" for 25mm and 37mm post-weighing
            If this option is selected, the user will be prompted to select either the header or the detail groups. The user will enter "1" for the header, and "2" for the detail group.
         C. "3" for Sentinel filter pre-weighing

If this option is selected, the user will be prompted to select either the header or the detail groups. The user will enter "1" for the header, and "2" for the detail group.

D. "4" for Sentinel filter post-weighing
If this option is selected, the user will be prompted to select either the header or the detail groups. The user will enter "1" for the header, and "2" for the detail group.
E. "5" for XRF analysis
F. "6" for Soil characterization
G. "7" Vacuum Dust characterization
H. "8" for 24-Hour Food Diary Check
If this option is selected, the user will be prompted to select either the header, the detail groups, the second page, or the supplement. The user will enter "1" for the header, "2" for the detail group, "3" for the second page, or "4" for the supplement.
I. "9" for Vacuum filter pre-weighing
If this option is selected, the user will be prompted to select either the header or the detail groups. The user will enter "1" for the header, and "2" for the detail group.
J. "0" to return to the previous menu

c. If a questionnaire is to be cleaned, the user must enter "3". The *MDM* will then prompt the user for a specific form. The user will then enter:

A. "1" for Descriptive Qx
B. "2" for Baseline Qx
C. "3" for Time Diary and Activity Qx
D. "4" for Follow Up Qx
E. "5" for Diet Diary Qx
F. "6" for Food Diary Follow-up
G. "7" for Technician Qx
H. "8" for Questionnaire Feedback Qx
I. "0" to return to the previous menu

4. If the specific form is the Descriptive Questionnaire then the user is required to enter three encryption keys that allow access to the Descriptive Questionnaire. If the encryption keys are improperly entered then a notice is displayed and the *MDM* ends.
5. The *MDM* shall prompt the user for an action. The user then selects an operation from a given list. Operations include:
   a. Create a new batch
   b. Clean existing batch

c. QA an existing batch
d. [Examine Descriptive Master Database]

6. If the user selects the Create a new batch option, the *MDM* will create a new batch and give notice of the new batch name. The user will then be prompted to press <enter> to continue program execution. The user is then asked if a batch list report is desired. If the user does desire such a report, a response of "Y" should be given (not case sensitive). If no report is desired, a response of "N" should be given. If this is the case, the *MDM* will list the report to the screen.
7. If the user selects the Clean existing batch option, the *MDM* will list all possible batches to clean. If there are no batches, the program will inform the user, and prompt the user to press the <enter> key, which will return the user to the previous menu. If there are batches to be cleaned, the user will be prompted for a specific batch. Once a batch is selected, it is preprocessed (checking for errors) for cleaning, and the user is given information regarding the status of the preprocessing.

    a. If no errors are found within in the data, the *MDM* will send mail to the Data Coordinator informing him of the newly cleaned batch.
    b. If errors are found in the data, the user is informed as to the number of errors found within the batch. For each error found in the batch, the user will be shown the key variables to locate the record containing the error, which is displayed between dashed lines. The user is then prompted for an action, the action being either the changing the erroneous value, the skipping of the error, a manual change to a different variable in the current record, a manual change to any variable in any record, or the user may quit.
        A. If the user wishes to change the erroneous value, then "C" must be entered. The variable name is then displayed, and the *MDM* prompts for a new value. Once a new value is given, the user is given the option to accept the value given, to accept a value formatted by the *MDM*, or to abort the operation. If the operation is not aborted, the user is prompted for a reason for the change. Once a reason is given, the update is completed.
        B. If the user wishes to skip the error, he/she must enter "S". This will cause the *MDM* to skip the current error, but not the current record.
        C. If the user wishes to change the value of a different variable in the current record, he/she must enter "M". The user will then be prompted for the number of changes to be made. Once this number is entered, the *MDM* will then prompt for the variable to be changed. Once a valid variable is specified, the user is prompted for the new

value. Once a new value is given, the user is given the option to accept the value given, to accept a value formatted by the *MDM*, or to abort the operation. If the operation is not aborted, the user is prompted for a reason for the change. Once a reason is given, the update is completed.

D. If the user wishes to change the value of a variable in a different record, he/she must enter "R". The user will then be prompted for the number of changes to be made. Once this number is entered, the *MDM* will then prompt for the key variable values of the record to be modified. Once valid values are given, the *MDM* prompts the user for the variable to be changed. Once a valid variable is specified, the user is prompted for the new value. Once a new value is given, the user is given the option to accept the value given, to accept a value formatted by the *MDM*, or to abort the operation. If the operation is not aborted, the user is prompted for a reason for the change. Once a reason is given, the update is completed.

E. If the user wishes to quit, he/she must hit the "Q" key. The user is then returned to the menu defined in 7.2.A.5.

8. If the user selects the QA an existing batch option, then the *MDM* will list the available batches to be QA'd. If no batches exist, the *MDM* will give notice to the user and return to the previous menu. If batches do exist, the user must specify which specific batch he/she wishes to QA. Once specified, the *MDM* will create a 10 percent file (or 50 percent file if the file had previously failed a 10 percent QA) by randomly selecting 10 percent of the records in the specified cleaning batch. This 10 percent file will identify the records to be used in the 10 percent QA (50 percent QA). For each record in the ten percent file (50 percent file), the *MDM* will cycle through the variables, displaying the values to the user. For each variable, the user will then have the following options:
   a. Entering "N" to accept the value as shown. The *MDM* will then display the next variable or record (as appropriate).
   b. Entering "C" to change the value of the variable. If this option is selected, the user is prompted for the new value. Once a new value is given, the user is given the option to accept the value given, to accept a value formatted by the *MDM*, or to abort the operation. If the operation is not aborted, the user is prompted for a reason for the change. Once a reason is given, the update is completed.
   c. Entering "Q" to quit. If the user quits without completing the current record, a notice is given that the entire record must be reviewed when QA is re-initiated on the batch.

d. When each of the selected records is reviewed, the *MDM* will return to the previous menu. Mail is sent to the Data Coordinator informing of the QA'd batch.

9. In the case where the user has selected the specific form of Descriptive Questionnaire the option to Examine the Descriptive database is available. If the user selects the Examine the Descriptive Database option, the *MDM* displays a new menu that gives the user the choice to Enter an Id or to Quit. If the user wants to quit he/she enters a value of "2" and is taken back to the previous menu. If the user enters a value of "1", he/she is prompted to enter an id for which to search. If an invalid id is entered an error message is displayed and the user is prompted to press the <enter> key to return to the previous menu. If the id entered was valid the key variables of the first occurrence of the record are displayed. The key variables displayed are the HHID, Date of questionnaire, Stage type, Contact number, and the first name of the person to whom the questionnaire was administered. At this point the user is prompted to continue search for updates of the specified questionnaire or to accept the current record. If continue is selected and no other records exist a message is displayed and the user is prompted to press <enter> to return to previous menu. If the current record is accepted then the user is prompted to enter variable names defined in the questionnaire dictionary to display the actual value for that variable. The user may quit at any time by entering the value 0, this returns the user to the previous menu.

B. Data Coordinator

1. Log into UNIX environment.
2. Type "data" to run the *MDM*.
3. Select form type.
   a. If a field form is desired, the user will enter "1". The *MDM* will then prompt the user for a specific form. The user will then enter:

      A. "1" for Household Summary Sheet
      B. "2" for PM Sampling
      C. "3" for PID Sampling
      D. "4" for Soil Sampling
      E. "5" for Floor Dust Sampling
      F. "6" for Surface Sampling
      G. "7" for Sentinel Sampling
      H. "8" for Personal Air Sampling
      I. "9" for Sample Tracking
         If this option is selected, the user will be prompted to select either the header or the detail groups. The user will enter "1" for the header, and "2" for the detail group.

J. "0" to return to the previous menu

b. If lab data is to be cleaned, the user will enter "2". The *MDM* will then prompt the user for a specific form. The user will then enter:

A. "1" for 25mm and 37mm pre-weighing
If this option is selected, the user will be prompted to select either the header or the detail groups. The user will enter "1" for the header, and "2" for the detail group.

B. "2" for 25mm and 37mm post-weighing
If this option is selected, the user will be prompted to select either the header or the detail groups. The user will enter "1" for the header, and "2" for the detail group.

C. "3" for Sentinel filter pre-weighing
If this option is selected, the user will be prompted to select either the header or the detail groups. The user will enter "1" for the header, and "2" for the detail group.

D. "4" for Sentinel filter post-weighing
If this option is selected, the user will be prompted to select either the header or the detail groups. The user will enter "1" for the header, and "2" for the detail group.

E. "5" for XRF analysis

F. "6" for Soil characterization

G. "7" Vacuum Dust characterization

H. "8" for 24-Hour Food Diary Check
If this option is selected, the user will be prompted to select either the header, the detail groups, the second page, or the supplement. The user will enter "1" for the header, "2" for the detail group, "3" for the second page, or "4" for the supplement.

I. "9" for Vacuum filter pre-weighing
If this option is selected, the user will be prompted to select either the header or the detail groups. The user will enter "1" for the header, and "2" for the detail group.

J. "0" to return to the previous menu

c. If a questionnaire is to be cleaned, the user must enter "3". The *MDM* will then prompt the user for a specific form. The user will then enter:

A. "1" for Descriptive Qx

B. "2" for Baseline Qx

C. "3" for Time Diary and Activity Qx

D. "4" for Follow Up Qx

E. "5" for Diet Diary Qx

F. "6" for Food Diary Follow-up
G. "7" for Technician Qx
H. "8" for Questionnaire Feedback Qx
I. "0" to return to the previous menu

4. If the specific form is the Descriptive Questionnaire then the user is required to enter three encryption keys that allow access to the Descriptive Questionnaire. If the encryption keys are improperly entered then a notice is displayed and the *MDM* ends.
5. The *MDM* will prompt the user for an action. The user then selects an operation from a given list. Operations include:
   a. Inspect cleaned batch
   b. Inspect QA'd batch
   c. [Examine Descriptive Master Database]
6. If the user selects the Inspect cleaned batch option, then the *MDM* will list the available batches to be inspected. If no batches exist, the *MDM* will give notice to the user and return to the previous menu. If batches do exist, the user must specify which specific batch he/she wishes to inspect. The user may then list all recommended changes to the screen (by pressing <L>), print out all recommended changes (by pressing <P>), accept all recommended changes (by pressing <A>, review changes individually (by pressing <R>), or quit (by pressing <Q>). If the user desires to review all the changes, for each recommended change, the *MDM* will list the type of change, the user requesting the change, the date the change was requested, the key variables for the record, the name of the corrected variable, the variable description, the old value, the suggested change, and the reason for the change. The user will then be given the option to accept the change, reject the change, skip the record, view the record again, or quit and continue later.
   a. If the user wishes to accept the change, then <A> must be pressed. The next recommended change will come up for review.
   b. If the user wishes to reject the change, then <R> must be pressed. The user is then prompted for a reason for the rejection. Once a reason is given, the next recommended change will come up for review.
   c. If the user wishes to skip the change, then <S> must be pressed. The next recommended change will come up for review.
   d. If the user wishes to view the record again, then <V> must be pressed. The current recommended change will then be displayed again.
   e. If the user wishes to quit and continue later, then <Q> must be pressed. The *MDM* will then display the previous menu.

Once all changes are reviewed, then if there were no rejections made, mail is sent to the appropriate data staff regarding the approval of the changes. The *MDM* then selects ten percent of the records in the batch for the Data

Technicians to QA. If rejections were made, then the data students are sent mail regarding the rejection of recommended changes, and the batch is returned for further cleaning if errors exist after the changes were rejected.

7. If the user selects the Inspect QA'd batch option, then the *MDM* will list the available batches to be inspected. If no batches exist, the *MDM* will give notice to the user and return to the previous menu. If batches do exist, the user must specify which specific batch he/she wishes to inspect. The user may then list all recommended changes to the screen (by pressing <L>), print out all recommended changes (by pressing <P>), accept all recommended changes (by pressing <A>, review changes individually (by pressing <R>), or quit (by pressing <Q>). If the user desires to review all the changes, then for each recommended change the *MDM* will list the type of change, the user requesting the change, the date the change was requested, the key variables for the record, the name of the corrected variable, the variable description, the old value, the suggested change, and the reason for the change. The user will then be given the option to accept the change, reject the change, skip the record, view the record again, or quit and continue later.
    a. If the user wishes to accept the change, then <A> must be pressed. The next recommended change will come up for review.
    b. If the user wishes to reject the change, then <R> must be pressed. The user is then prompted for a reason for the rejection. Once a reason is given, the next recommended change will come up for review.
    c. If the user wishes to skip the change, then <S> must be pressed. The next recommended change will come up for review.
    d. If the user wishes to view the record again, then <V> must be pressed. The current recommended change will then be displayed again.
    e. If the user wishes to quit and continue later, then <Q> must be pressed. The *MDM* will then display the previous menu.

   Once all changes are reviewed, then if there were no rejections made, mail is sent to the appropriate data staff regarding the approval of the changes. The number of corrections are then counted, and if the number of corrections is less than 5 percent of the total data fields, the data is considered to be cleaned. Otherwise:
    f. If more than 5 percent of the total fields were changed, and the batch represents a ten percent QA, then the *MDM* will select fifty percent of the records in the original batch for a more thorough QA.
    g. If more than 5 percent of the total fields were changed and the batch represents a fifty percent QA, the batch is destroyed and the data students are sent notice of the need to re-scan the batch.

If rejections were made, then the data students are sent mail regarding the rejection of recommended changes, and the batch is returned for further QA.

8. In the case where the user has selected the specific form of Descriptive Questionnaire the option to Examine the Descriptive database is available. If the user selects the Examine the Descriptive database option, the *MDM* displays a new menu that gives the user the choice to Enter an Id or to Quit. If the user wants to quit he/she enters a value of 2 and is taken back to the previous menu. If the user enters a value of 1 he/she is prompted to enter an id for which to search. If an invalid id is entered an error message is displayed and the user is prompted to press the <enter> key to return to the previous menu. If the id entered was valid the key variables of the first occurrence of the record are displayed. The key variables displayed are the HHID, Date of questionnaire, Stage type, Contact number, and the first name of the person to whom the questionnaire was administered. At this point the user is prompted to continue search for updates of the specified questionnaire or to accept the current record. If continue is selected and no other records exist a message is displayed and the user is prompted to press <enter> to return to previous menu. If the current record is accepted then the user is prompted to enter variable names defined in the questionnaire dictionary to display the actual value for that variable. The user may quit at any time by entering the value 0, this returns the user to the previous menu.

C. Data Manager

1. Log into UNIX environment.
2. Type "data" to run the *MDM*.
3. Select form type.
   a. If a field form is desired, the user will enter "1". The *MDM* will then prompt the user for a specific form. The user will then enter:

      A. "1" for Household Summary Sheet
      B. "2" for PM Sampling
      C. "3" for PID Sampling
      D. "4" for Soil Sampling
      E. "5" for Floor Dust Sampling
      F. "6" for Surface Sampling
      G. "7" for Sentinel Sampling
      H. "8" for Personal Air Sampling
      I. "9" for Sample Tracking
         If this option is selected, the user will be prompted to select either the header or the detail groups. The user will enter "1" for the header, and "2" for the detail group.

J. "0" to return to the previous menu

b. If lab data is to be cleaned, the user will enter "2". The *MDM* will then prompt the user for a specific form. The user will then enter:

A. "1" for 25mm and 37mm pre-weighing
If this option is selected, the user will be prompted to select either the header or the detail groups. The user will enter "1" for the header, and "2" for the detail group.

B. "2" for 25mm and 37mm post-weighing
If this option is selected, the user will be prompted to select either the header or the detail groups. The user will enter "1" for the header, and "2" for the detail group.

C. "3" for Sentinel filter pre-weighing
If this option is selected, the user will be prompted to select either the header or the detail groups. The user will enter "1" for the header, and "2" for the detail group.

D. "4" for Sentinel filter post-weighing
If this option is selected, the user will be prompted to select either the header or the detail groups. The user will enter "1" for the header, and "2" for the detail group.

E. "5" for XRF analysis

F. "6" for Soil characterization

G. "7" Vacuum Dust characterization

H. "8" for 24-Hour Food Diary Check
If this option is selected, the user will be prompted to select either the header, the detail groups, the second page, or the supplement. The user will enter "1" for the header, "2" for the detail group, "3" for the second page, or "4" for the supplement.

I. "9" for Vacuum filter pre-weighing
If this option is selected, the user will be prompted to select either the header or the detail groups. The user will enter "1" for the header, and "2" for the detail group.

J. "0" to return to the previous menu

c. If a questionnaire is to be cleaned, the user must enter "3". The *MDM* will then prompt the user for a specific form. The user will then enter:

A. "1" for Descriptive Qx

B. "2" for Baseline Qx

C. "3" for Time Diary and Activity Qx

D. "4" for Follow Up Qx

E. "5" for Diet Diary Qx

F. "6" for Food Diary Follow-up
G. "7" for Technician Qx
H. "8" for Questionnaire Feedback Qx
I. "0" to return to the previous menu

4. If the specific form is the Descriptive Questionnaire then the user is required to enter three encryption keys that allow access to the Descriptive Questionnaire. If the encryption keys are improperly entered then a notice is displayed and the *MDM* ends.
5. The *MDM* shall prompt the user for an action. The user then selects an operation from a given list. Operations include:
   a. Append Data
   b. File Integrity Check
   c. Global Change
   d. Other functions
   e. Quit
6. If the user selects the Append Data option, then the *MDM* will list the available batches to be appended. If no batches exist, the *MDM* will give notice to the user and return to the previous menu. If batches do exist, the user must specify which specific batch he/she wishes to append. Program proceeds to append the data and mail out notice to data manager and data analyst.
7. If the user selects the File Integrity Check option, the *MDM* marks all the new records in the master database as having completed a file integrity check.
8. If the user selects the Global Change option, the *MDM* prompts user to perform a <G>lobal change or to <Q>uit. If user selects global change then program prompts user to enter the name of the variable they wish to make a global change on, and the *MDM* will continue to ask until a valid variable is entered. Once a valid variable is entered, the user is prompted for key variable names and corresponding values in order to properly identify target records. Once all key variables have been defined then user is allowed to enter the new global change value. Once a global value is given, the user is given the option to accept the value given, to accept a value formatted by the *MDM*, or to abort the operation. If the user does not abort a second prompt is given which asks if the user wishes to continue with global change. If the user proceeds with global change, he/she is prompted for a reason for the change. Once a reason is given, the update is completed.
9. If the user selects the Other Function option, the *MDM* displays another menu of commands used primarily by other data personnel. The options (and procedures) include:
   a. Create New Batch
      (see 7.2.A.6)
   b. Clean Existing Batch

(see 7.2.A.7)

c. QA Existing Batch
(see 7.2.A.8)

d. Inspect Cleaned Batch
(see 7.2.B.6)

e. Inspect QA'd Batch
(see 7.2.B.7)

f. Univariate Check
(see 7.2.D.6)

g. Bivariate Check
(see 7.2.D.7)

h. Quit

D. Data Analyst

1. Log into UNIX environment.
2. Type "data" to run the *MDM*.
3. Select form type.
   a. If a field form is desired, the user will enter "1". The *MDM* will then prompt the user for a specific form. The user will then enter:

      A. "1" for Household Summary Sheet
      B. "2" for PM Sampling
      C. "3" for PID Sampling
      D. "4" for Soil Sampling
      E. "5" for Floor Dust Sampling
      F. "6" for Surface Sampling
      G. "7" for Sentinel Sampling
      H. "8" for Personal Air Sampling
      I. "9" for Sample Tracking
      If this option is selected, the user will be prompted to select either the header or the detail groups. The user will enter "1" for the header, and "2" for the detail group.
      J. "0" to return to the previous menu

   b. If lab data is to be cleaned, the user will enter "2". The *MDM* will then prompt the user for a specific form. The user will then enter:

      A. "1" for 25mm and 37mm pre-weighing
      If this option is selected, the user will be prompted to select either the header or the detail groups. The user will enter "1" for the header, and "2" for the detail group.
      B. "2" for 25mm and 37mm post-weighing

If this option is selected, the user will be prompted to select either the header or the detail groups. The user will enter "1" for the header, and "2" for the detail group.

C. "3" for Sentinel filter pre-weighing
If this option is selected, the user will be prompted to select either the header or the detail groups. The user will enter "1" for the header, and "2" for the detail group.

D. "4" for Sentinel filter post-weighing
If this option is selected, the user will be prompted to select either the header or the detail groups. The user will enter "1" for the header, and "2" for the detail group.

E. "5" for XRF analysis

F. "6" for Soil characterization

G. "7" Vacuum Dust characterization

H. "8" for 24-Hour Food Diary Check
If this option is selected, the user will be prompted to select either the header, the detail groups, the second page, or the supplement. The user will enter "1" for the header, "2" for the detail group, "3" for the second page, or "4" for the supplement.

I. "9" for Vacuum filter pre-weighing
If this option is selected, the user will be prompted to select either the header or the detail groups. The user will enter "1" for the header, and "2" for the detail group.

J. "0" to return to the previous menu

c. If a questionnaire is to be cleaned, the user must enter "3". The *MDM* will then prompt the user for a specific form. The user will then enter:

A. "1" for Descriptive Qx
B. "2" for Baseline Qx
C. "3" for Time Diary and Activity Qx
D. "4" for Follow Up Qx
E. "5" for Diet Diary Qx
F. "6" for Food Diary Follow-up
G. "7" for Technician Qx
H. "8" for Questionnaire Feedback Qx
I. "0" to return to the previous menu

4. If the specific form is the Descriptive Questionnaire then the user is required to enter three encryption keys that allow access to the Descriptive Questionnaire. If the encryption keys are improperly entered then a notice is displayed and the *MDM* ends.
5. The *MDM* shall prompt the user for an action. The user then selects an operation from a given list. Operations include:

6. Univariate check
7. Bivariate check
8. If the user selects the Univariate check option, the *MDM* marks all the new records in the master database as having completed a univariate check.
9. If the user selects the Bivariate check option, the *MDM* marks all the records that have not previously passed a bivariate check.

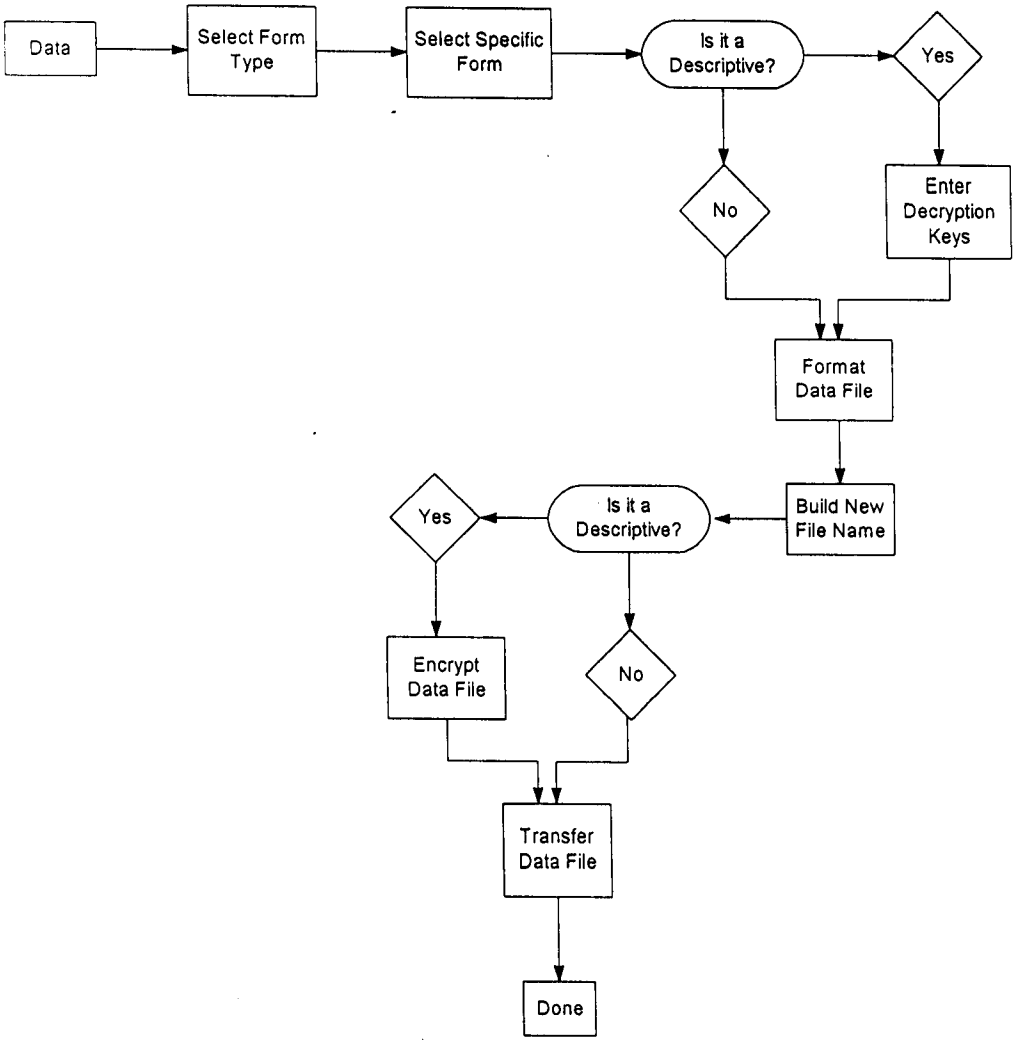
## 8.0 Records

Any changes made to a batch are recorded in its corresponding Change Log. This log records the author of the change, the date the change was made, the variable changed, the value the variable was changed from, the value the variable was changed to, the reason for the change, the data the change was reviewed, whether or nor the change was accepted or rejected, and the reason for the acceptance or rejection of the change. The Change Log is always stored in the same directory that the cleaning batch is.

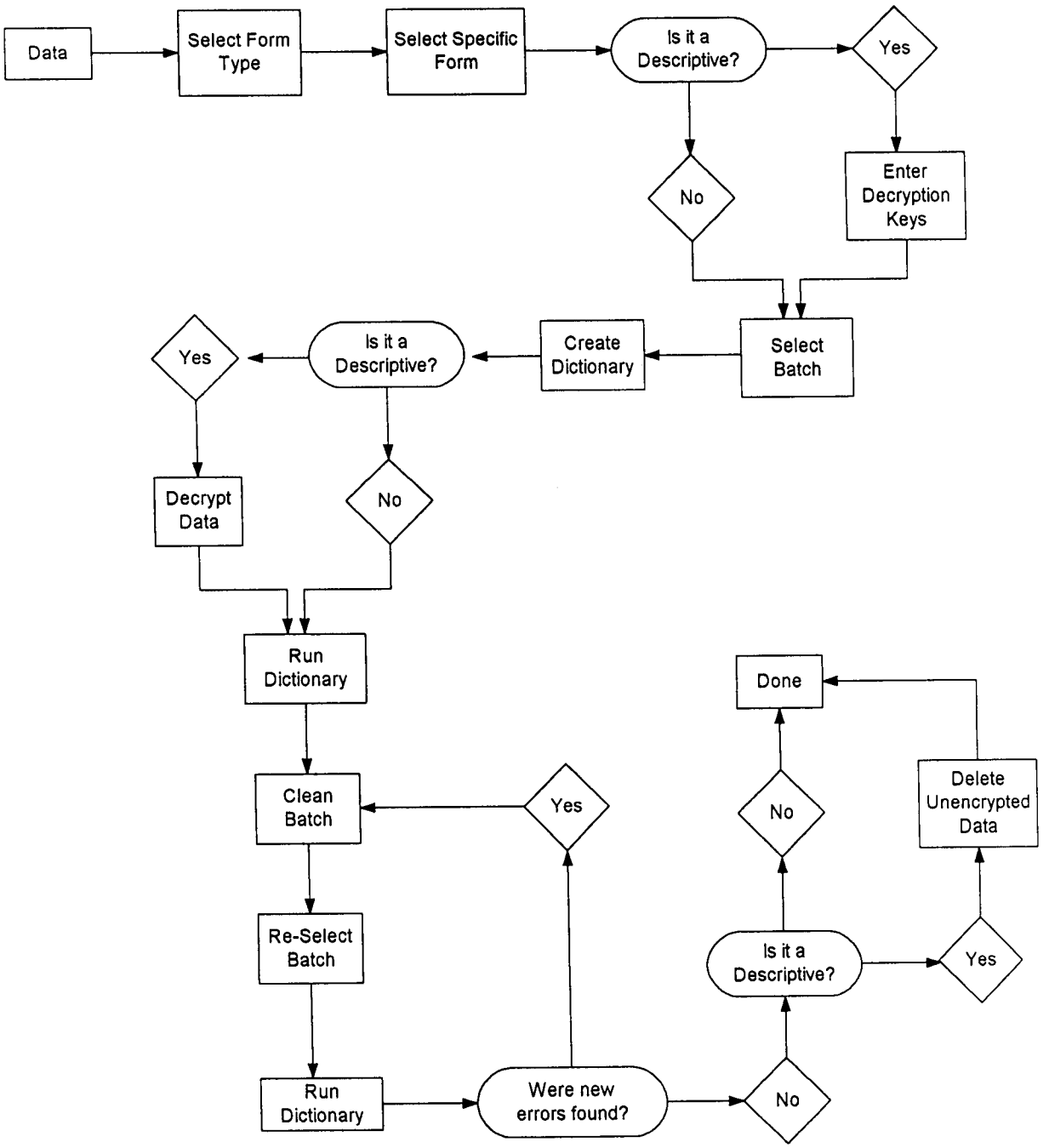
**Inclusions:**

Appendix A: Batch Creation (1 page)
Appendix B: Cleaning of Batches (1 page)
Appendix C: QA'ing of Batch (1 page)
Appendix D: Inspecting of a Cleaned Batch (1 page)
Appendix E: Inspecting of a QA'd Batch (1 page)
Appendix F: Appending of Data (1 page)
Appendix G: File Integrity Check (1 page)
Appendix H: Univariate Check (1 page)
Appendix I: Bivariate Check (1 page)
Appendix J: Performing a Global Change (1 page)
Appendix K: Changing of Encryption Keys (1 page)

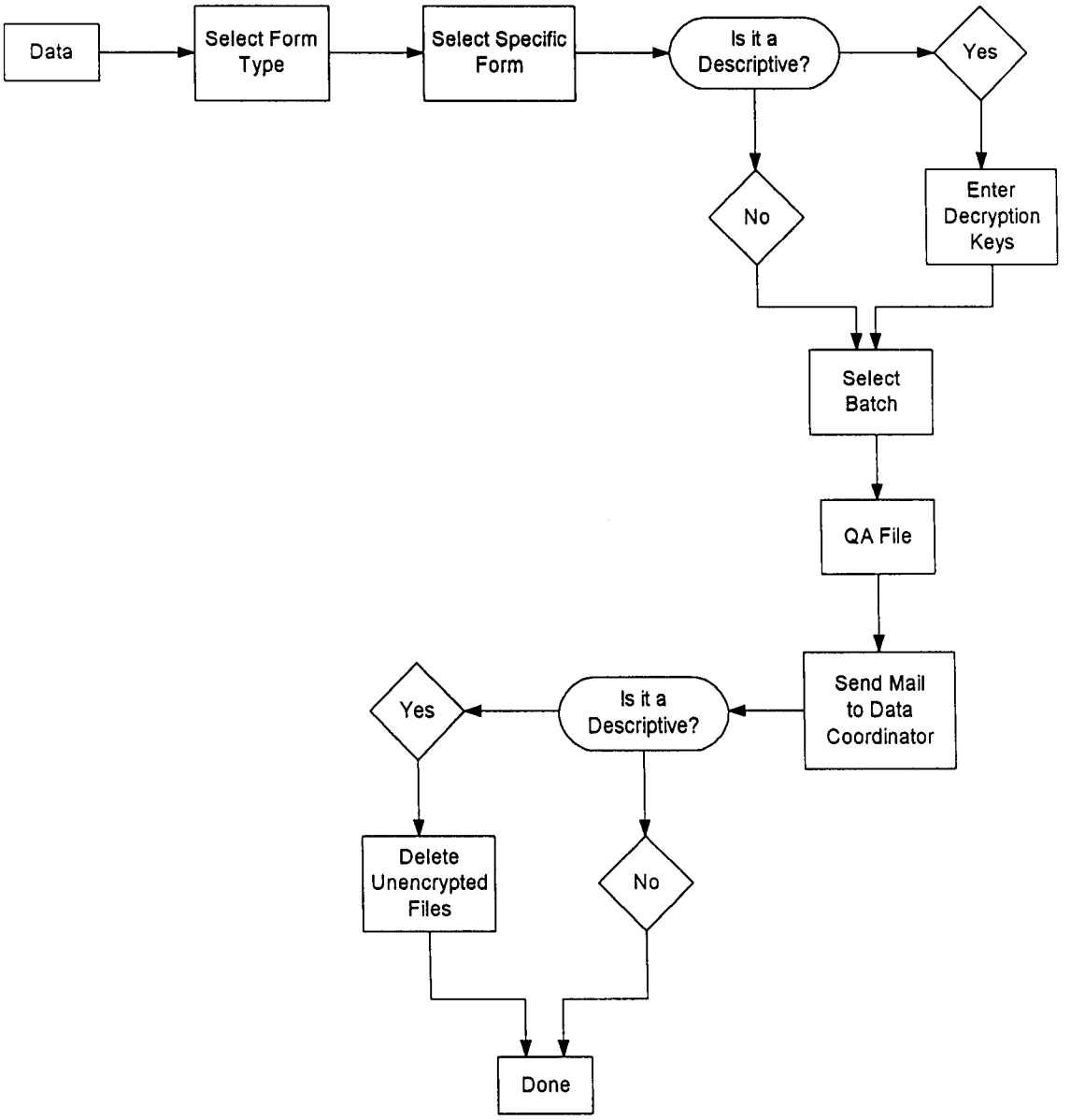
# Appendix A: Batch Creation



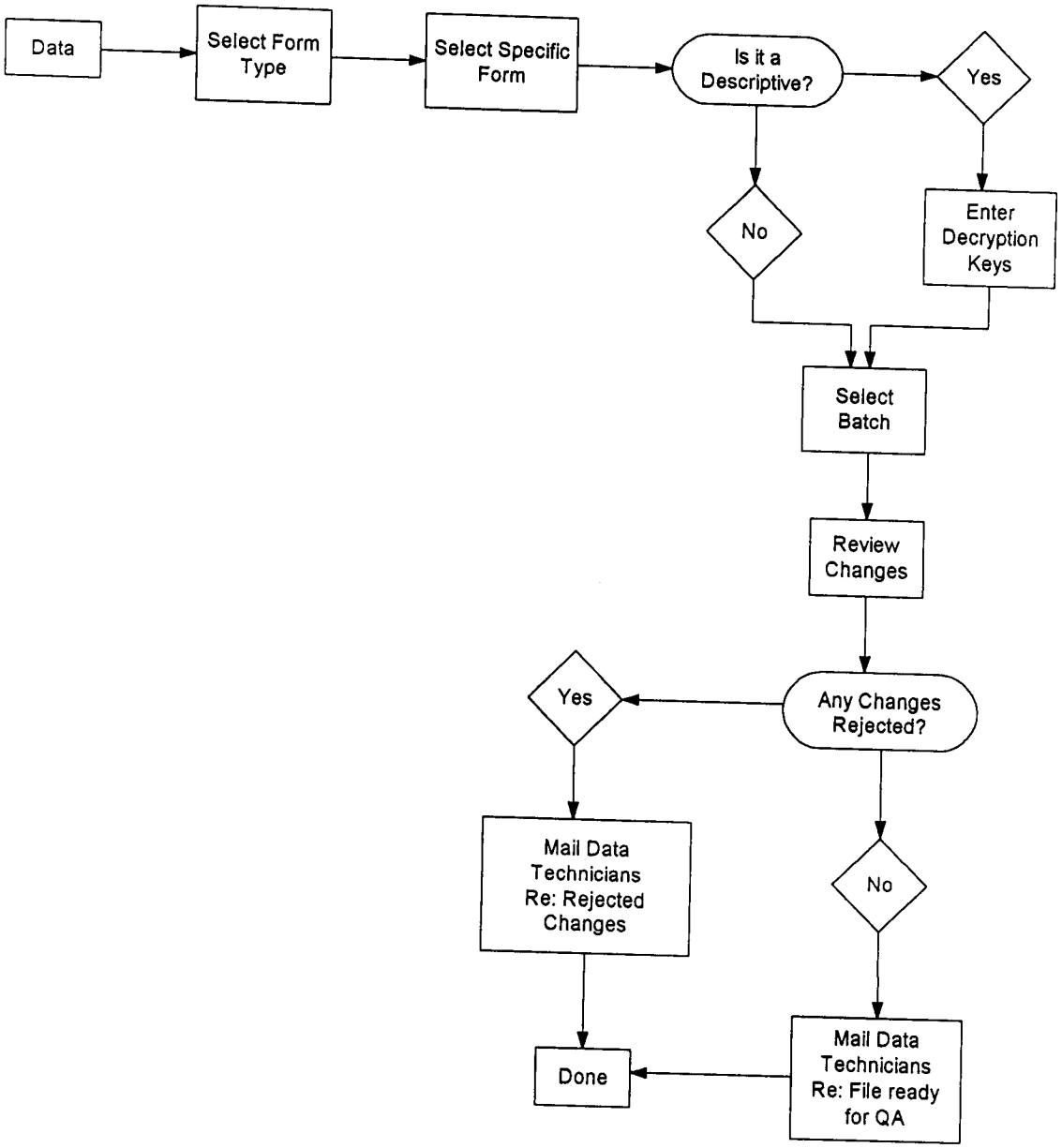
## Appendix B: Cleaning of Batches



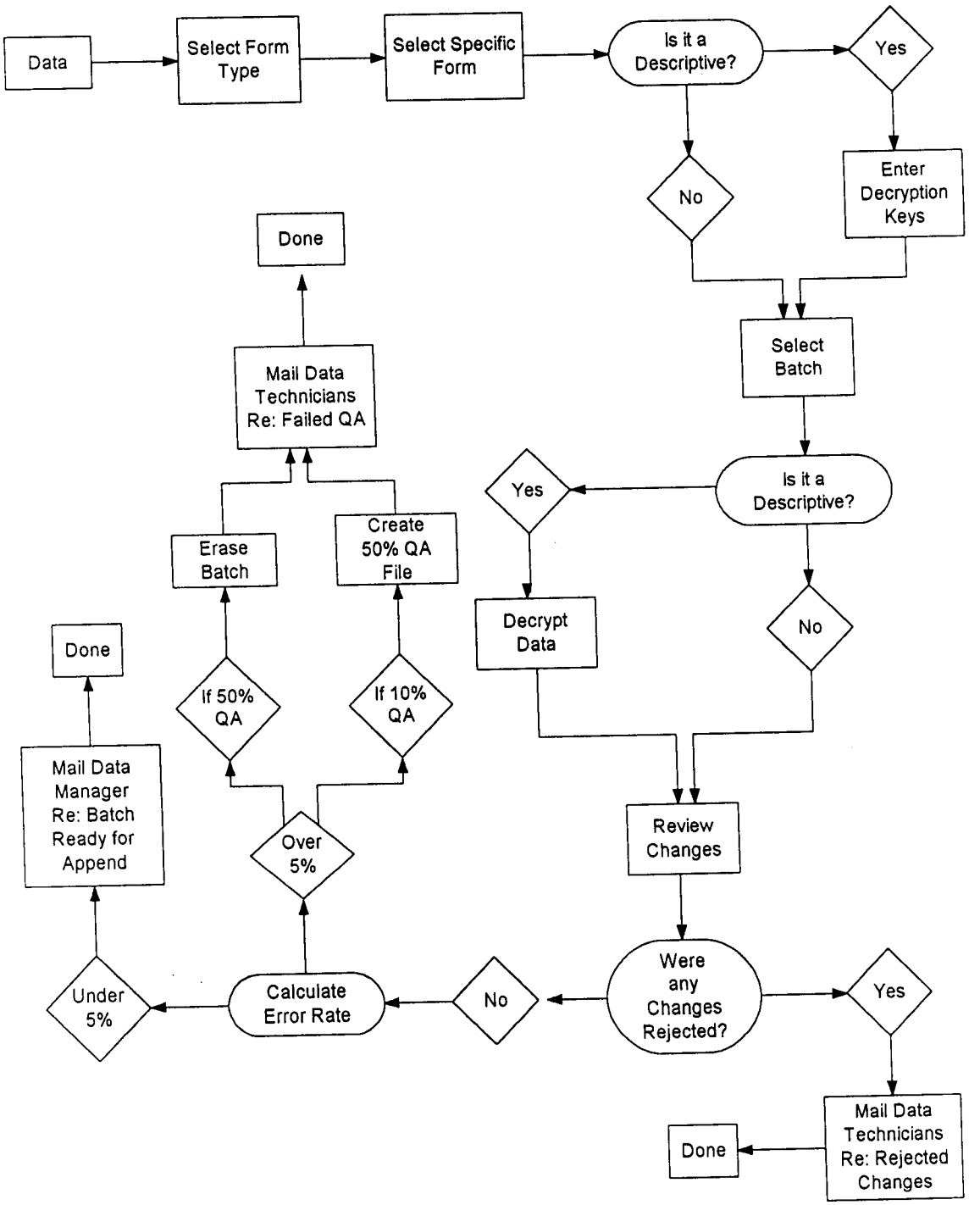
# Appendix C: QA'ing of Batch



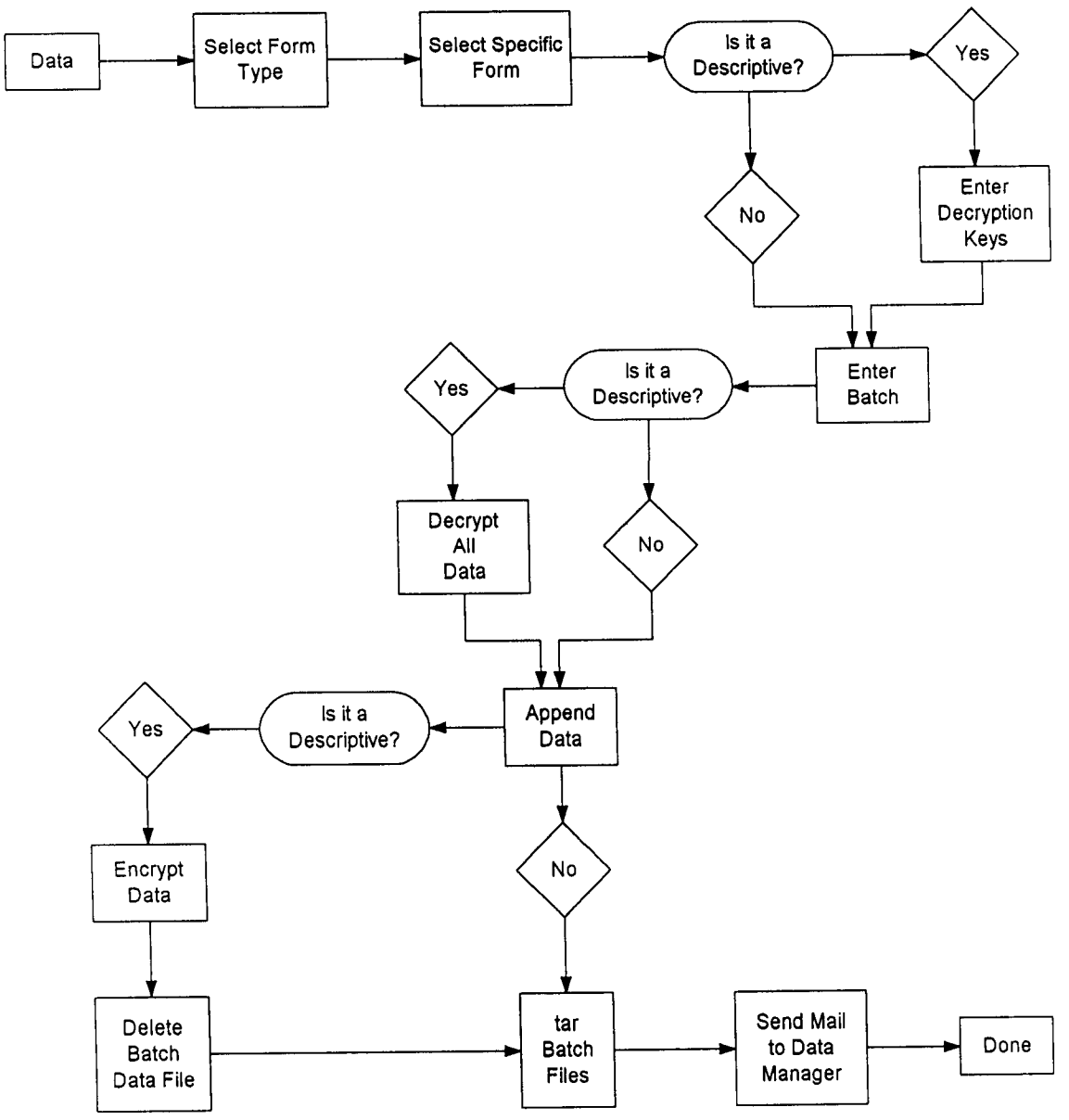
## Appendix D: Inspecting of a Cleaned Batch



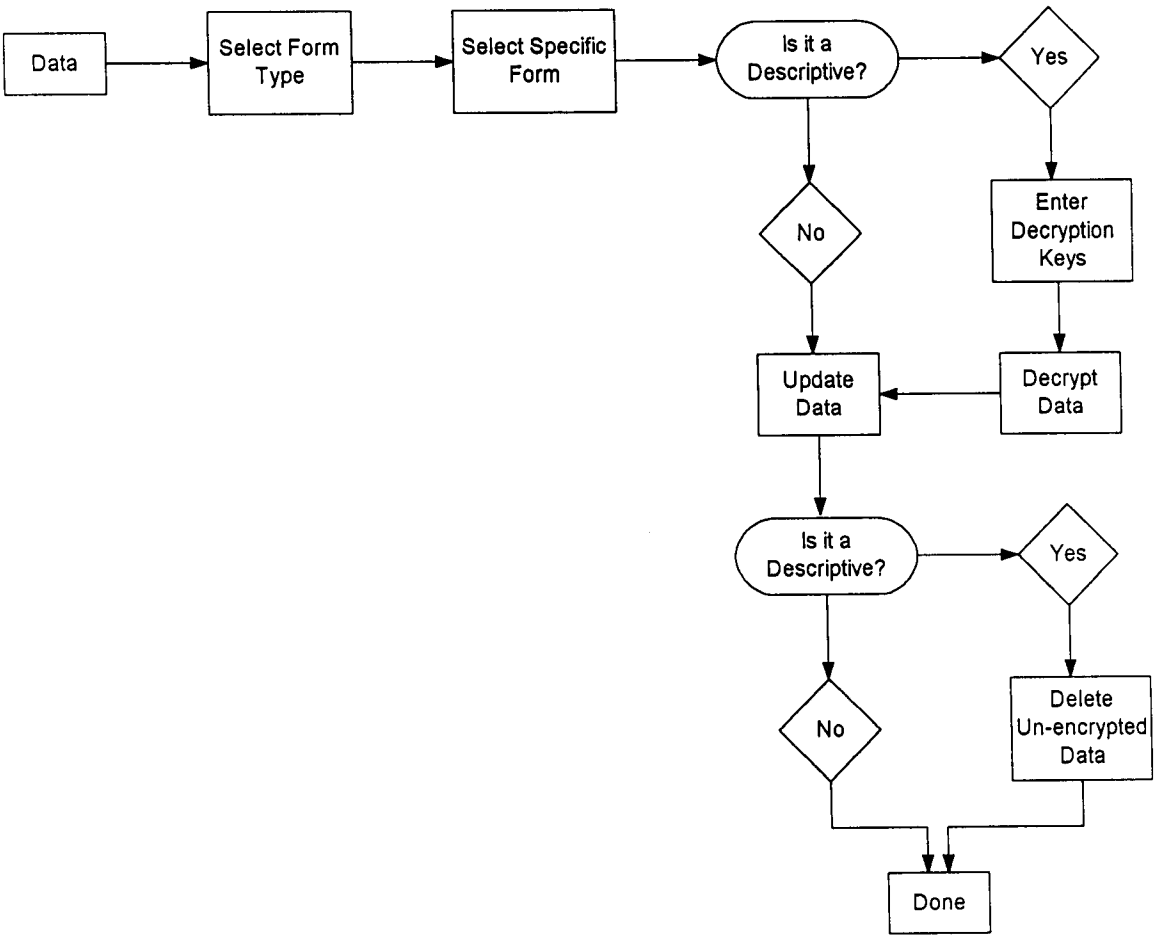
# Appendix E: Inspecting of a QA'd Batch



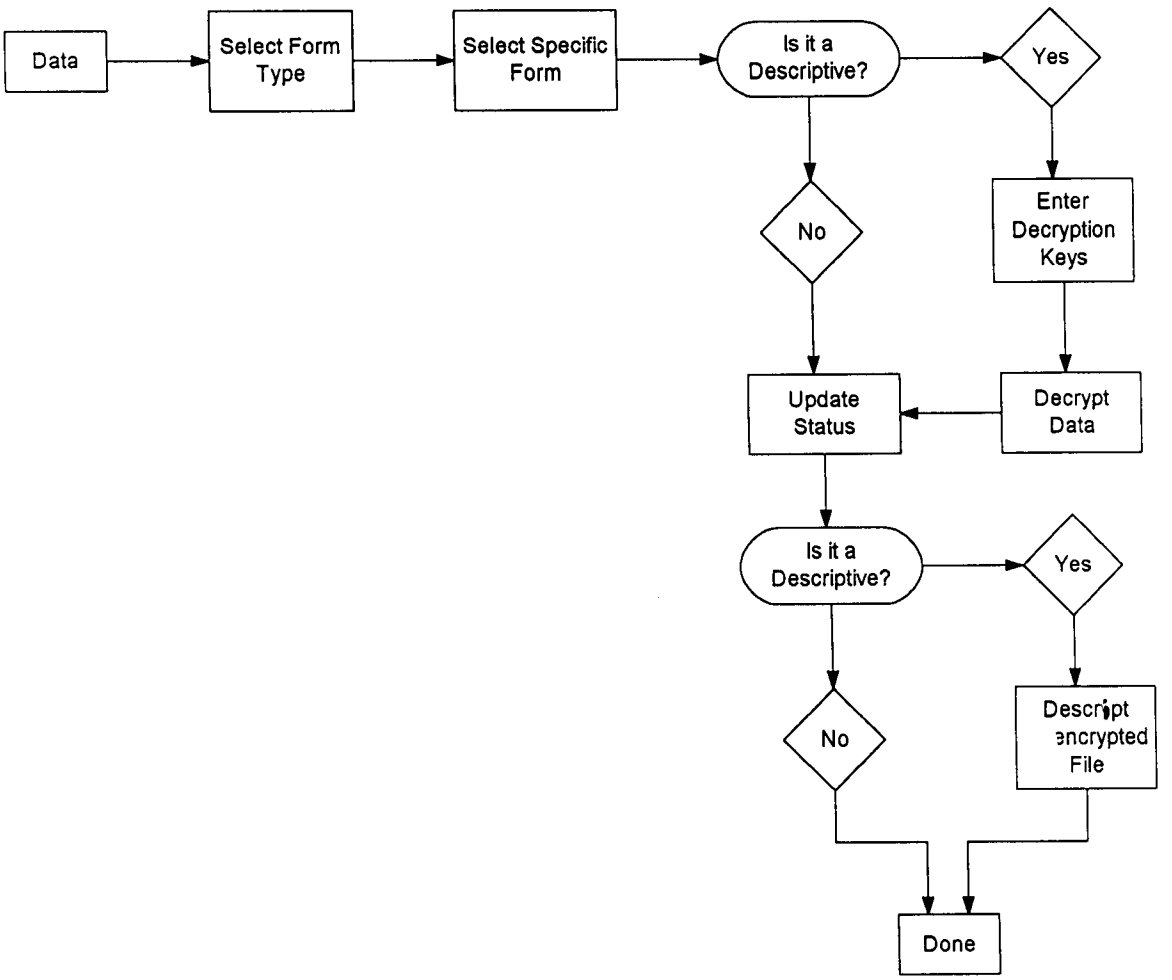
# Appendix F: Appending of Data



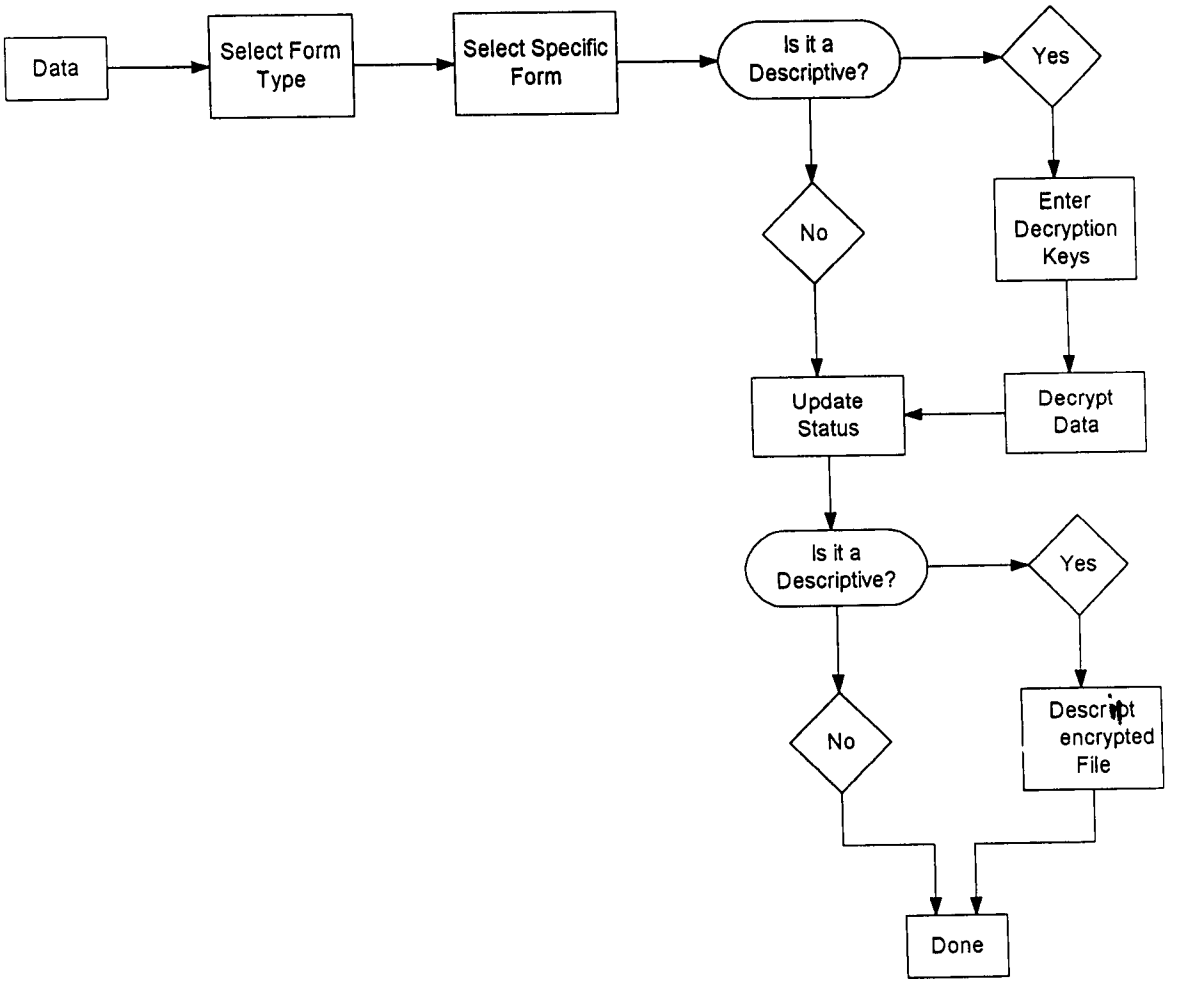
# Appendix G: File Integrity Check



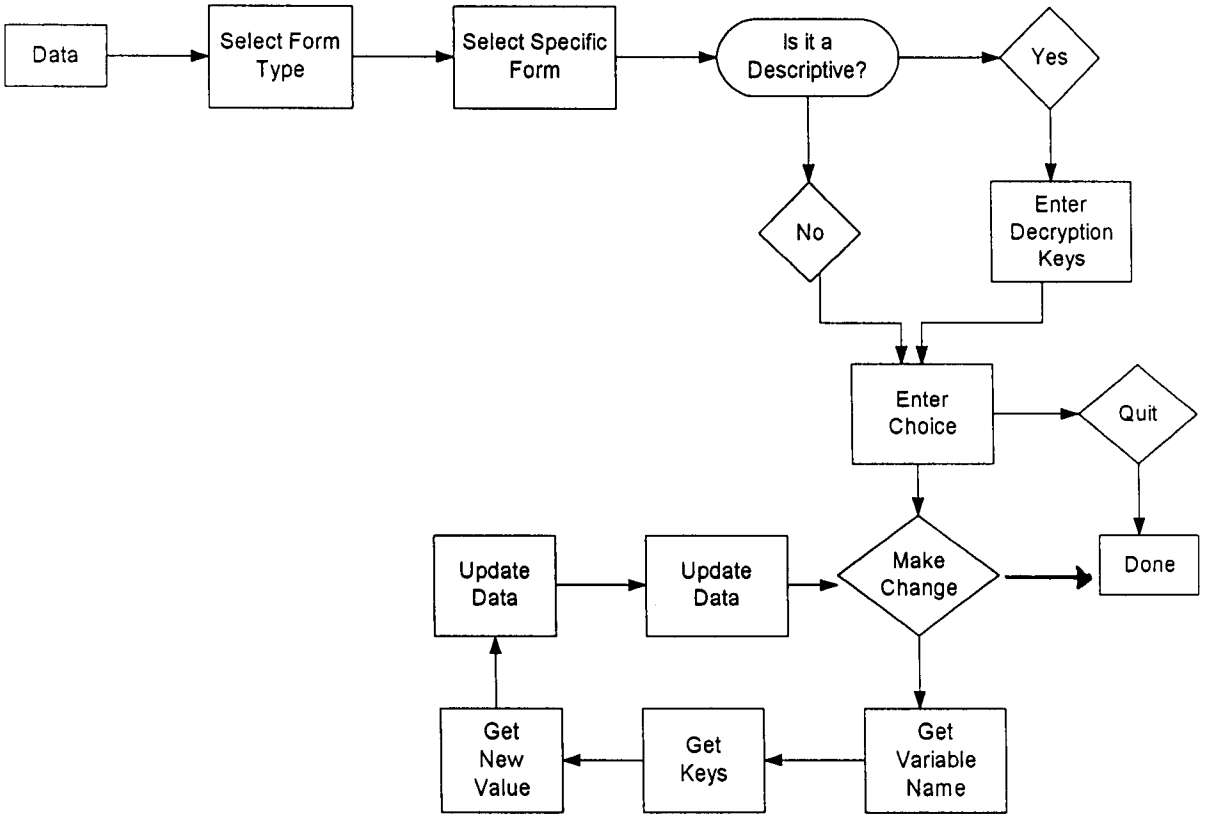
## Appendix H: Univariate Check



## Appendix I: Bivariate Check



# Appendix J: Performing a Global Change



## Appendix K: Changing of Encryption Keys