

National Human Exposure Assessment Survey (NHEXAS)

Arizona Study

Quality Systems and Implementation Plan for Human Exposure Assessment

The University of Arizona
Tucson, Arizona 85721

Cooperative Agreement CR 821560

Standard Operating Procedure

SOP-UA-D-27.0

Title: Instructions for the Addition of Individual Cleaned Non Scanned
Data Batches to Master Databases

Source: The University of Arizona

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

Notice: *The U.S. Environmental Protection Agency (EPA), through its Office of Research and Development (ORD), partially funded and collaborated in the research described here. This protocol is part of the Quality Systems Implementation Plan (QSIP) that was reviewed by the EPA and approved for use in this demonstration/scoping study. Mention of trade names or commercial products does not constitute endorsement or recommendation by EPA for use.*

**Instructions for the Addition of
Individual Cleaned Non-Scanned Data
Batches to Master Databases**

1.0 Purpose and Applicability

This SOP describes the procedures to be used for appending cleaned individual data batches to the master databases. This procedure applies to the NHEXAS study, 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 CLEANING: 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."
- 2.3 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.
- 2.4 DATA, ELECTRONIC: Data stored on some type of magnetic or optical medium (e.g., floppy disk, hard disk).
- 2.5 DATA, PHYSICAL: A datum or data written on a physical data form.
- 2.6 DATA PROCESSING BATCH: A number or alphanumeric code assigned to either a collection of household packets, or to a collection of one or more site or lab sheets within a site or lab packet. It is abbreviated as DP Batch#.
- 2.7 DATA PROCESSING ERROR: An error occurring 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.
- 2.8 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.

- 2.9 DATABASE, MASTER: This is the accumulative database generated from validated data cleaning batches. Newly cleaned batches are appended to the master database.
- 2.10 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.
- 2.11 FORM, PHYSICAL = The paper or "hard copy" version of a data form. This is also referred to as a "physical data form."
- 2.12 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.
- 2.13 HRP SITE: The **H**ealth **R**elated **P**rofessions building, located at 1435 North Fremont Avenue, Tucson, AZ 85719. This is an annex of the Arizona Prevention Center and the primary site of NHEXAS Arizona.
- 2.14 NHEXAS Arizona: Acronym for **N**ational **H**uman **E**Xposure **A**ssessment **S**urvey, a research project conducted in Arizona by the University of Arizona/Battelle/Illinois Institute of Technology consortia.

3.0 References

SOP# UA-D-25.X

SOP# UA-D-44.X

4.0 Discussion

This SOP describes the appendage of non-scannable data to master databases. The appendage of all scanned data is covered in SOP# UA-D-44.X. Non-scannable data is data obtained from Battelle, the Environmental Protection Agency (EPA), or the Centers for Disease Control (CDC).

5.0 Responsibilities

The Project Data Manager is responsible for appending all data to the master databases.

6.0 Materials and Reagents

6.1 H&E Local Area Network

6.2 For Data from the EPA

6.2.1 Data Diskette

6.2.2 Physical Data Descriptors

7.0 Procedure

7.1 Appending NHEXAS Scannable Forms to the Master Databases

See SOP# UA-D-44.X

7.2 Appending NHEXAS XRF Download Data to the Master Database

7.2.1 All download data files are stored in /rsc53/NHEXAZdata/xrf_dat

7.2.2 From this directory, log on as super-user and run *transfer* (by simply typing "transfer"). This will archive all files, as well as move them to the appropriate in-process directory.

7.2.3 Change directory to /rsc53/NHEXAZdata/master_process/xrf_data. From there, the *generate* program must be run (by simply typing "generate").

7.2.4 Do a listing of the files in
/rsc53/NHEXAZdata/master_process/xrf_data.
There will be one or more files with names of the form master_X.dat, where X is some number. For each filename of this type, go to the directory /rsc53/NHEXAZdata/master/labaz/xrf/download/X.

7.2.5 From each of these directories, run the *xrf update* program (by simply typing "xrf_update"). This program will archive the current master database, and then update it.

7.3 Appending data obtained from the E.P.A. to the Master Databases (i.e., food and water data)

7.3.1 Data batches are received in the mail from a postal delivery service addressed to the Data Manager. These batches contain hard copy analysis results, along with the results in files on diskettes.

7.3.2 The Data Manager records the date of reception on both the hard copy and the diskettes.

7.3.3 The hard copy are filed in a file cabinet in order of date.

7.3.4 The files on each diskette are run through the E.P.A. data formatting program "Standard Format Translator User Interface, V1.0b", which results in ASCII files. Then, the deciphered files are archived in /rsc53/NHEXAZdata/inproces/labepa/output, and the original files in /rsc53/NHEXAZdata/inproces/labepa/input.

7.3.5 Change directory to:
/rsc53/NHEXAZdata/inproces/labepa/progs/stage1.

7.3.6 Run the program runme (by simply typing "runme"). This program will generate a file called newresult.dat. This file is copied to /rsc53/NHEXAZdata/master/epa_data/epa_master.dat. This file is the new master database.

7.4 Appending data obtained from Battelle to the Master Databases (i.e., food and water data)

7.4.1 The most current versions of the food and water data sets are FTP'd by Battelle personnel onto the HRP workstation Lonicera. These data sets, which are saved in Microsoft Access data file formats, are stored in
/rsc53/NHEXAZdata/inproces/labbat.

7.4.2 All master data files can be created by exporting the tables from these Microsoft Access files.

7.4.3 These master data files will be stored in
/rsc53/NHEXAZdata/master/labbat.

7.5 Appending data obtained from the Centers for Disease Control to the Master Databases

This process does not yet exist. Once the transfer methods from CDC are finalized, this SOP will be updated to reflect the final process.

8.0 Records

Records for updates of EPA data are stored in a file cabinet in the Data Manager's office at the HRP site. These records include documentation of receipt of data (which is logged on the physical form accompanying all data), and the date of appendage (which is written as a date on the disk containing the data). Records for updates of Battelle data are logged in a file which resides in /rsc54/brinkman. Once transfer methods from CDC are finalized, this SOP will be updated to report the appropriate records.