DRAFT Documentation for the data files for the National Hazardous Waste Constituent Survey

Prepared by Westat, Inc.
Revised March 30, 1998

1. INTRODUCTION

The National Hazardous Waste Constituent Survey (NHWCS) was designed and conducted by the US Environmental Protection Agency (EPA) Office of Solid Waste. The survey collected information about the physical characteristics and chemical constituents of hazardous waste streams. The information will be used primarily to support EPA's Hazardous Waste Identification Rule (HWIR). The population that received the survey included the largest treatment, disposal, and recycling facilities in the United States. Since EPA did not mandate that the facilities complete the survey, participation was voluntary. Despite the voluntary nature of the survey, a 71% response rate was achieved. The survey was designed and conducted by ICF and Westat under the direction of EPA.

The survey data were entered into electronic data files for use. This report describes the contents of the data files as of the date shown above. The data have been carefully reviewed. Respondents were contacted by phone to review and correct inconsistent responses. As of the date above, all respondents for whom there were questions had been called. After reviewing the inconsistent responses with the respondents and obtaining corrections, revisions were made in the data files. The data files contain the revisions for all but 11 respondents. Repeated attempts were made to contact these respondents, however, their inconsistent responses remained unresolved at the time the data review was stopped.

The topics described in the documentation are shown below.

2. SURVEY DESIGN STEPS

2a. Draft questionnaire

EPA, ICF and Westat designed a draft data collection instrument and instructions to be used for the pretest and the Information Collection Request for OMB. The design anticipated that waste stream specific data from the sites' 1993 Hazardous Waste Report would be printed on the survey form.

2b. OMB Information Collection Request (ICR)

Westat assisted ICF in the preparation of the EPA's ICR with specific attention to the sampling plan for facilities and the waste streams of interest within facilities. The Information Collect Request was submitted to OMB by EPA and approved.

2c. Pretest

Nine facilities were selected to participate in a test (or pretest) of the questionnaire and survey procedures prior to conducting the main survey. The pretest used the initial draft of the NHWCS. Facilities were selected to give the widest possible range of volume and types of wastes that were managed, as well as geographic distribution of the sites.

The pretest facilities were contacted (or debriefed) to determine how the survey questionnaire and procedures might be improved. A debrief was conducted with all but one of the facilities. Seven of the facilities were debriefed by conference telephone call and there was one facility visit. The objective of the debrief was to determine whether respondents understood the intent of the survey and how the data would be used, whether the instructions and questions were clear or ambiguous, and whether the data were readily available. EPA was also interested in respondent burden. Several of the debrief questions were designed to elicit the staff cost and time required to complete the survey, the type and level of staff that were involved in completing the survey, and whether the survey could be restructured to reduce response time.

2d. Revised Questionnaire

The data collection instrument and instructions were revised to incorporate the feedback from the pretest respondents. Specifically, the questions were reordered and reformatted to fit onto fewer pages so that the overall size of the survey would be less daunting and a few questions were reworded to be more succinct.

2e. Implementation of the Full Survey

The NHWCS questionnaire was sent to 212 facilities. Each of the facilities received a survey with preprinted information on selected hazardous waste streams that were managed by the facility in 1993. The 1993 Hazardous Waste Report National Oversight Database was used as the source for facility selection and for waste stream information. The preprinted facility and waste stream information is included in the data files and referred to as the preloaded data. For each of the selected waste streams, respondents were asked to provide information on the physical characteristics (such as specific gravity, percent oil and grease by weight, flash point, or BTU content) and constituent concentrations (on a whole waste or leachate basis). In order to minimize the burden for the largest of these facilities, the maximum number of waste streams for each facility was limited to 20.

3. QUESTIONNAIRE FORMAT

The survey had two parts. The first, Part I, comprising one page, contained facility information such as the facility name and address, EPA ID, and the name and phone number of the person completing the survey. The facility name, location and mailing address, EPA ID, and the name of the person completing the 1993 Hazardous Waste Report were obtained from the BRS system and printed (preloaded) on the questionnaire. Respondents were asked to verify and correct errors in the preloaded data.

Part II had information about the waste streams. Part II was divided into two sections. The first section included two pages for each waste stream covered by the survey. The number of waste streams covered by the survey ranged from 1 to 20 per facility. The waste stream information covered the quantity and source of the waste, waste codes and waste description, system types in which the waste stream was managed, and physical characteristics of the waste. Information on the waste codes, waste description, waste form code, system types, and quantities was preloaded from the BRS system. Respondents were asked to correct, if necessary, the preloaded information and to enter information on the physical characteristics of the waste stream.

The second section of Part II requested information about the average concentration of constituents in the waste stream. Constituents that were expected to be present, based on the waste codes, were preloaded in the questionnaire. Respondents were asked to add other constituents that were present. If the average concentration could not be determined, the respondent was asked to make an estimate.

4. HOW THE SAMPLE WAS SELECTED

The survey was sent to hazardous waste treatment, disposal, and recycling facilities (TDRs) that manage the largest waste stream volumes, as reported in the 1993 Hazardous Waste Report (also known as the 1993 Biennial Report), within one or more of the following categories of waste: total waste, listed waste, characteristic waste, mixed listed and characteristic waste, non-wastewaters, and combusted waste. The largest facilities, which together handled over 90% of the wastes (by volume), in these six categories, were selected for the survey. The facilities were not randomly selected from a larger population.

The survey requested information about major waste streams within the selected facilities. For the purposes of this survey, a major waste stream is defined as being greater than 400 tons for non-wastewaters or 40,000 tons for wastewaters. If there were more than 20 major waste streams, 20 waste streams were randomly selected, with larger waste streams having a higher probability of selection. If the facility managed several "like" streams (with the same origin codes, form codes, and waste codes) the smaller "like" streams were aggregated for the purposes of identifying major waste streams. If the aggregated stream was greater than 400 tons for non-wastewaters or 40,000 tons for wastewaters, it was considered a major stream. For each aggregated stream, one individual stream was selected to represent the aggregated stream.

The data files contain two variables that contain statistical weights. One variable is used to scale the quantities managed for the smaller waste streams up to the quantity for the aggregated stream. The second is used to scale up the quantity for the major or aggregated waste stream up to the quantity handled by the facility when there are more than 20 major waste streams.

5. SURVEY OPERATIONS

5a. History of Survey Distribution and Retrieval

The surveys were sent via Federal Express to 212 facilities on August 9, 1996. The package contained a memo from Michael Shapiro of EPA, a copy of the survey with preprinted information for each selected waste stream, and a set of instructions with attached code lists. A separate page was inserted behind the cover memo providing the due date of the survey (September 13, 1996) and the return address. A blank page and instructions for adding additional constituents were also inserted. The telephone help line number was printed in the instructions and at the top of each page of the survey.

On August 30, the 200 facilities who had not responded were sent, via Federal Express, a second copy of the survey and instructions, a memo from Michael Shapiro dated August 19, a separate sheet with the September 13 due date and return address, constituent addendum instructions and blank sheet, a list of unit of measure codes (omitted from the original instructions), and special instructions for Texas respondents alerting them to possible problems with facility-specific data.

As of the September 13 due date, fewer than half of the selected facilities had completed and returned their surveys. EPA agreed to extend the survey period by another month. A week after the first due date, on September 23, a second follow-up contact was made with almost 90 facilities by Federal Express that included a cover memo from Michael Shapiro, and a separate sheet with the return address and new extended deadline of October 18.

After discussion with EPA, ICF and Westat extended the survey period beyond October 18, 1996 to accept surveys until December 8, 1996. During this extension period EPA contacted some facilities to encourage participation in the survey.

Six responses were received in the two months after the end of the survey period.

5b. The NHWCS Telephone Help Line

The US EPA established a toll free telephone help line to assist respondents with completion of their National Hazardous Waste Constituent Surveys. Almost 180 calls were received and logged in during the data collection period between August 12 and October 18, 1996 and the short extended period at the end of the survey period. The help line operated from 8:00 a. m. to 5:00 p.m. Eastern Time. A voice mail system recorded messages from respondents who called during hours that the help line was not staffed.

A Telephone Help Line Log was prepared and delivered October 31, 1996, for EPA that listed each of the questions and responses. The introduction to the Telephone Log summarizes the types of questions that were received during the help line operations.

The toll free number was also used during data retrieval when respondents responded to specific questions or missing data elements on the completed surveys.

6. RESPONSE RATE

In all, 147 facilities in the full survey plus 9 pretest facilities returned completed surveys (71% complete with n = 221 facilities, 212 survey facilities and 9 pretest facilities). Of the 65 facilities not responding to the survey, eight facilities (3.8%) called and specifically declined to participate. Nine facilities (4%) called the help line or wrote to say that due to resource

constraints such as reduced staffing and changes in facility status (no longer operating as a RCRA waste management facility, for example) they would not complete the survey. An additional survey was filed with EPA as proprietary Confidential Business Information (CBI). The data from this survey are not included in the data files.

EPA received constituent data for large waste streams from four Rollins Corporation facilities in electronic format. These date were received in association with facility visits made by EPA. The Rollins data are provided in separate data files documented in Appendix 5.

Response rates can be calculated in different ways. The variable WSDATA on the facility file contains "Yes" for each of the 156 facilities that provided some constituent characteristic or concentration data. Several facilities provided only characteristic information for waste streams or provided constituent information, however only for waste streams that were not managed on-site. Only 152 facilities provided constituent data for waste streams that are of primary interest to the survey objectives, waste streams managed on-site. The variable HASCCONC on the facility file has a "Y" for each of the 152 facilities which provided constituent concentration data for waste streams managed on-site. The variables WSDAT and HASCCON indicate waste streams with some constituent concentration data and with constituent concentration data for waste streams managed on-site.

7. DATA ENTRY

7a. Data Entry of Survey Responses

The completed surveys were processed for key entry in Westat's proprietary CO-ED system. Data supplied by the respondents were appended to the preloaded (preprinted) information. In instances where respondents corrected the preprinted information, these data elements were updated. Part of Westat's standard key entry procedure is to double key enter every data element. A special program is then used to compare both sets of keyed data and any discrepancies trigger an error report. The reports are reviewed by a supervisor who made the final determination about the correct data item. All completed NHWCS were double key entered and verified, including the pretest surveys.

7b. Key Entry of Pretest Data

The format and order of questions in the survey questionnaire were different than in the pretest questionnaire. Therefore, the questionnaires from the pretest respondents were transcribed from the pretest forms on to copies of the survey questionnaire for data entry.

One question was added to the survey questionnaire that was not asked in the pretest. This is item B3, Was this waste stream managed exclusively in units exempt from RCRA permitting requirements? In the transcription from pretest to full survey forms, this item was left blank.

8. QUALITY REVIEW PROCEDURES

8a. Reviewing and Revising the Data

The data were reviewed by staff who were familiar with the BRS reporting requirements, the survey and waste processing technology. When confusing or inconsistent responses were

identified, the respondent was contacted to clarify how the wastes were handled and how the processing and constituent data were reported. Revisions were entered into spreadsheet files for later inclusion into the data files.

8b. Advanced Quality Control

After the data were key entered and reviewed, a decision was made that the basic quality control procedures originally envisioned would not be sufficient. The basic procedures involved examining each variable in isolation to ensure that the entries were consistently coded and making sure that the concentrations and physical characteristics were within reasonable ranges. The basic checks would not capture inconsistencies between variables.

The advanced quality control approach taken for the NHWCS involved looking at the entire submission, including waste codes, system types, form code, SIC code, origin code, physical characteristics, chemical constituents, and constituent concentrations to decide if the data were internally consistent and valid. Each facility's data were reviewed manually and potential inconsistencies were flagged. The facilities were contacted and the submission reviewed.

Of the 221 facilities that received the NHWCS (212 facilities plus the 9 pretest facilities), 156 facilities responded (147 facilities plus the 9 pretest facilities). Data from all 156 facilities were thoroughly reviewed. Two facilities were not contacted, as their submission appeared to be error-free. Of the remaining 154 facilities that were contacted, data for 143 facilities were discussed with the respondents and corrections made. The remaining 11 facilities were called multiple times but the submissions remained unresolved at the time the QA/QC process was stopped. The 11 facilities are identified in Appendix 1. None of these 11 facilities were a pretest facility.

8c. Item-by-Item Analysis

PRELOADED DATA

The survey sent to the respondents contained preloaded information from the 1993 Biennial Reporting System (BRS). This data was used in also designing the sample frame. The respondents were asked to review the preloaded information and make changes as necessary.

Section A of the survey contains facility information such as address and contact name. This section was not analyzed for this exercise. Section B contains waste stream quantities and waste characteristics. Questions B1, B4, B5, B6, B7, B8, B9, B10, B11, B12, B13 were preloaded from the 1993 BRS database. Questions B2 and B3 were also preloaded but the data did not come directly from the BRS database. Question B2 refers to whether a waste stream was a wastewater. This variable was determined using the form code, system type and the description of the waste. Question B3 refers to whether the waste was managed in units subject to RCRA permitting. This data element was computed using information from Forms PS and IC of the Biennial Report. Questions B14, B15, B16, and B17 were blank and the respondents were requested to provide the information.

Section C of the survey refers to the constituents concentration and the section was preloaded with CAS numbers and names of chemicals likely to be present in the waste based on the EPA waste codes.

FINDINGS

Appendix 2 contains the response/change rate of each question present in Sections B and C of the survey. For each question in Section B, the table reveals the number of preloaded responses that were changed (in cases where the data were preloaded) or provided (in cases where the data were not preloaded), not changed, added (in cases where the data element was blank in the 1993 BRS database or the respondent added a waste stream), or deleted (in cases where the preloaded data element was deleted or the waste stream was deleted). Beneath the response numbers are the response rates. The rates were calculated by dividing the number of responses by the total number of waste streams.

For example, B1 - Waste form code - was a preloaded data element. Of the 1,167 waste streams in the database, 210 or 18 percent of the form codes were missing in the preloaded data and these codes were obtained during the QA/QC process, 892 or 76 percent of the form codes were confirmed to be correct and did not change from the preloaded data, 59 or 5 percent of the waste streams had an erroneous waste form code in the preloaded data and the data element was changed, and 6 waste form codes were deleted due to the deletion of waste streams that contained a preloaded waste form code.

Another example, B16_1 - Specific gravity - was not a preloaded data element. Of the 1,167 waste streams, the respondents provided the specific gravity for 960 or 82 percent of them, while for 207 waste streams, specific gravity was not reported.

The most dramatic change was made to Questions B3. The waste streams selected for this survey were wastes that were managed in units subject to RCRA permitting. Wastes managed in units exempt from RCRA permitting would not be affected by the proposed Hazardous Waste Identification Rule (HWIR) and hence were omitted from the survey. However, 49 (4 percent) waste streams were identified during the advanced QA/QC process as being managed in units exempt from RCRA permitting (see response rate to Question B3). These waste streams account for 45 percent of the total quantity of waste for which responses were received. This is very significant considering that the number should be close to zero. The error was not caused by the sampling procedure, but rather by the incomplete/incorrect data present in the 1993 BRS database.

Questions 14, 15, and 16 had no preloaded data and the response rate for these questions was pretty high. When the characteristic was not applicable to a particular waste stream, the response was left blank. This explains the low response rate for characteristics such as total ash content, flash point, % total suspended solids (TSS), % oil and grease, 5-day biological oxygen demand (BOD), and 5-day chemical oxygen demand (COD). If the property was applicable for a particular management method and was not entered, the respondent was called for that information. For example, only 56 percent of all waste streams have a BTU content present but this number is closer to 100 percent for all waste streams that were combusted.

The respondents were asked to provide the concentration of constituents on a whole waste and leachate basis. This information was collected in Question C1. Of the 1,167 waste streams in the survey, there were 215 (18 percent) for which both concentrations were provided, 765 responses (66 percent) had only the whole waste concentration, 88 responses (8 percent) had only the leachate concentration, and 99 responses (8 percent) had no concentration present.

8d. General Observations

When the respondents were called to discuss their data submission, most of the changes made

were to the preloaded data obtained from the 1993 BRS database and relatively few changes were made to the new information collected on the survey such as Question C1, constituent concentration data.

The Biennial Report forms are designed to collect information on a system basis. That is, if a waste was stabilized prior to landfill, then the correct reporting would be to indicate the waste being managed in a stabilization system and the resultant waste (different quantity and characteristics) being landfilled. At least two States, Ohio and South Carolina, advise their waste handlers to report only the final treatment or disposal of the waste. In the above example, the waste managers in these two States would report only landfill and there would be no information on whether the waste was stabilized prior to landfill. Moreover, the amount landfilled would be incorrect if the waste was stabilized since the quantity increases when stabilization is done.

A few generators suggested that EPA devise an F code for incinerator ash similar to the F039 code for leachate. The present regulations require that the waste handler assign the listed waste codes of all wastes combusted to the ash and in some cases, the ash is assigned upwards of fifty waste codes.

9. DATA PROCESSING PROCEDURES

9a. Making CAS Numbers Consistent

The respondents were asked to provide constituent names and CAS numbers for constituents that were present but not in the preloaded questionnaire. In cases where the respondents added constituents, CAS numbers were often not provided. When constituent names were provided, they were sometimes misspelled and sometimes it was not clear which constituent was being identified. When CAS numbers were present they were sometimes incorrect (such as two digits were reversed). In other cases, constituents were listed which did not have a CAS number, such as "Soil and debris."

The efforts to make CAS numbers consistent required comparing the respondent information with available lists of CAS number and constituent names. In many cases there are multiple names for the same constituent. There may also be several CAS numbers that apply to a constituent, for example, when the constituent has several isomers. Each isomer would have its unique CAS number and a separate CAS number might exist for unspecified mixtures of the isomers. When in doubt, the CAS number provided by the respondent was used.

Valid CAS number-constituent name combinations were obtained from or verified on the CambridgeSoft Corporation web site (http://chemfinder.camsoft.com/). Chemists at Westat and EPA reviewed the names to select one name as the primary or preferred name for the NHWCS data files. Commonly used alternate names were also included in the list of valid CAS-constituent name pairs. Based on a formula available on the web site, the validity of the CAS numbers in the database was checked (the last digit is a check digit). All CAS numbers in the CAS variable are valid according to the formula.

In order to revise the CAS numbers, three files were prepared. The first file contained a master list of CAS number - constituent name pairs that were considered valid for the survey. The file included all recognized chemicals reported as constituents in the NHWCS. This file also included constituents (such as "Soil and debris") for which no CAS number exists. For these constituents the CAS number is "No CASRN." For constituent names that were misspelled, the second file had the incorrect spelling paired with the correct CAS number and correct constituent

name. The third file had corrections for the CAS numbers, listing the respondent's CAS number and constituent name and the corresponding corrected CAS number and constituent name. The constituent name was assumed to be misspelled when the constituent name provided by the respondent was not valid but was similar to a valid constituent name, and the CAS number was either not valid or matched the CAS number of the valid name. CAS number corrections were made when the respondent's CAS number could not have been valid (such as it contained letters) or had an obvious typo. In addition, the CAS numbers for metals and metal compounds were set to the CAS number for the metal. Also, two forms of cyanide were given the CAS number for cyanide with an extra letter appended to the end (57-12-5A for Cyanides (Amenable) and 57-12-5C for Cyanides (Cyanohydrins)).

These files were used to match the CAS number and constituent name pairs provided by the respondent with the correct CAS number and constituent pairs. The files were first matched on both CAS number and constituent name, then on CAS number and finally on constituent name. The match was performed after removing punctuation characters (comma, dash, and space). Thus names with different punctuation and the same spelling would match. Following guidance from EPA, when the CAS number and constituent names were both valid but were inconsistent, the CAS number was assumed to be correct. Appropriate matches were found for almost all constituents. The data files contain the respondent CAS number and constituent pairs and the corrected CAS number and constituent pairs.

9b. Weights

The file contains several weight variables. Each of these weight variables is described below.

INSURWGT

A few respondents provided information on waste streams which were not part of the survey. These waste streams were generally provided because the respondent thought they might be affected by future HWIR regulations. There is a weight called INSURWGT which is 1 if the waste stream was one of the waste streams originally requested as part of the survey and 0 if the waste stream was not requested.

In some cases, the respondent concluded that the requested waste stream was better represented, for purposes of the survey, as two separate waste streams. In this case they provided separate information on physical characteristics and constituents for two waste streams which, when combined, corresponded to the one waste stream requested in the survey. Both waste streams are in the data files with the same statistical weights. The variable WSTYPE (waste stream type) indicates which waste streams are splits of the originally requested waste stream. For example, if waste stream 1 was reported as two waste streams, WSTYPE = 'Split 1A' for one of the two splits and 'Split 1B' for the other. Both of the waste streams have INSURWGT = 1 to indicate that the waste streams requested in the survey covered both splits.

BRSWGT

In some cases there were waste streams selected for the survey that were incorrectly reported in the BRS. Possible reasons for the incorrect BRS information might include: (1) the respondent incorrectly reported information to their state and later corrected the error, however the correction was not entered into BRS, (2) as a result of conversations with the respondents, they realized that the original report was not correct, or (3) state processing of the data added records compared to what the respondent thought was correct. In any case, these waste streams were either duplicates or should not have been reported in BRS, according to the respondents. The variable BRSWGT contains 0 if the waste stream was incorrectly reported in BRS, otherwise it

contains 1.

AGGWGT

In the selection of waste streams, some smaller waste streams with similar characteristics were aggregated into a larger aggregated waste stream. If the aggregated waste stream was over the volume cutoff and was selected for the survey, one of the individual waste streams which makes up the aggregate was selected with probability proportional to size (volume). The variable AGGWGT contains weights to scale up from the one selected waste stream to the total for all similar aggregated waste streams. If AGGWGT = 1 then the waste stream is not an aggregated stream. If AGGWGT is greater than one, the waste stream is a representative waste stream from an aggregate waste stream and AGGWGT is the ratio of the volume of the aggregate waste stream to the volume of the selected waste stream. There were a few cases in the pretest facilities where the waste streams were not sampled as planned and the same waste stream was selected more than once. In these cases, AGGWGT = 1 for the first waste stream and 0 for the duplicate waste streams.

SAMPWGT

If there were more than 20 waste streams larger than the volume cutoff, a random sample of 20 was selected with probability proportional to size. The variable SAMPWGT contains either a 1, indicating that the waste was selected with certainty, or a value greater than 1, indicating that the waste stream was selected with probability proportional to the volume of the waste stream. When sampling with probability proportional to size and selecting multiple waste streams, some waste streams may be so large that they will be selected for sure, or with certainty (that is with probability equal to 1.0). After removing waste streams selected with certainty, the remaining waste streams are selected with probabilities proportional to their size.

In order to select which waste streams are included in an analysis:

- Use INSURWGT if you want to eliminate waste streams that were volunteered by the respondents.

- Use BRSWGT to eliminate waste streams that were incorrectly reported and should not have been in the BRS.

In order to adjust for the sampling:

- No weights are needed to report on only the waste streams in the survey.

- Use AGGWGT to scale up from the requested individual waste streams to the aggregated waste streams.

- Use SAMPWGT and AGWGT to scale up from the sampled waste streams to all waste streams at a facility which, after aggregation, are greater than the selected cutoff.

Multiply the appropriate weights to get the final weight. Multiply the final weight by the quantities to get the quantities adjusted for the sampling procedures. For example, to calculate the volume of waste represented by the survey which should have been reported to BRS, calculate the weighted sum for volume, with

WEIGHT1 = INSURWGT*BRSWGT* SAMPWGT *AGGWGT.

To calculate the number of waste streams (both those above the cutoff and those below which when aggregated are above the cutoff), sum WEIGHT1. To calculate the number of "major"

waste streams represented by the data, use WEIGHT2 = INSURWGT*BRSWGT* SAMPWGT.

9c. Reporting of Not Tested (NT) and zero concentrations

The respondents may have reported NT, indicating that the constituent was not tested, ND, indicating that the constituent was tested for but was below the detection limit, or zero, indicating either a very low concentration or that the constituent was not present.

EPA was interested in whether the constituent was present. Present is a relative term; present at what level? "Present" was interpreted to mean present at any level, no matter how small, which might be of interest to EPA for regulatory purposes. The respondent was asked to make the judgment as to whether the constituent was present. Respondents were contacted to clarify whether zero meant not present (NA or not applicable to the waste), was present and not tested for (PR), or present and tested for but less than the detection limit (ND). When the constituent was not tested for, the respondents were asked if the constituent was present (PR) or not present (NA). If the respondent was unsure if the constituent was present or absent, the response was coded as DK for Don't Know.

In order to minimize the respondent burden, if the respondent provided concentrations for the whole waste but not the leachate test or the leachate but not the whole waste test, we assumed the leachate or whole waste test, respectively, was not performed. We did not call the respondent to get information for tests that we assumed were not performed. In order to implement this procedure, if all constituents were labeled as NT for a test, we assumed that the test was not performed. Also, if data were provided for both tests and the constituent was reported as present for one test, we did not call the respondents to review NT responses for the same constituent in the other test, unless the respondent was called to review other responses. In these cases, very few respondents were able to provide information on the test with all NT responses.

We were not able to contact a few respondents to clarify the meaning of the not tested responses and the zeros. Therefore, some of these responses are still in the data files.

Based on the telephone follow-up, if an assumption must be made about the zero and NT concentrations, we suggest that they be interpreted as meaning that the constituent is not present (NA).

9d. Obtaining SIC codes for Off-site Waste Generation

The SIC code printed on the survey (preloaded data) in Question B5 pertains to the site that generated the waste (in the variable SICCODE). Since the Survey was only sent to sites that managed hazardous waste, the SIC code was verified or collected (if the preloaded data were missing) only for those wastes that were generated and managed at the same site. Since most of the wastes in the Survey were not generated at the site they were managed, the SIC code was blank for the majority of the wastes.

To obtain the SIC code of the generators that did not receive the Survey but shipped the waste to generators in the Survey, and for those generators in the Survey that did not provide a SIC code, programming was done to extract the SIC code from the following data bases (FINDS/RCRIS, 1993 BRS, 1991 BRS and 1995 BRS) sequentially until a SIC code was obtained. The site SIC code is present on Form GM in the 1993 and 1995 BRS databases, but on Form IC in the 1991 BRS database. If the site was not present in these four databases or the site failed to report a SIC code, then the final SIC code was left blank. Sometimes, the preloaded SIC code on the Survey

form or FINDS/RCRIS was not very useful (e.g., "9999" - Nonclassifiable establishment) or invalid ("2800"). For these cases, a more meaningful SIC code was sought from the three BRS databases in the same sequence as indicated before. The revised SIC codes for the generators of the waste, whether on-site or off-site, and in the variable SIC.

10. SPECIAL FACTORS AFFECTING THE USE AND INTERPRETATION OF THE DATA

Not all respondents completed the information for all waste streams for several reasons. In some cases facilities did not have constituent data for all waste streams. Some facilities did not complete information for waste streams that they felt would not be affected by HWIR. Several other factors are detailed below.

10a. State Reporting Differences

State differences in reporting 1993 Hazardous Waste data to the EPA's Biennial Reporting System (BRS) National Oversight Database had an impact on the level of effort required to complete the NHWCS. Four states - Texas, South Carolina, Tennessee, Oregon, representing over one fourth of the facilities in the NHWCS - use different reporting systems for collecting data that is subsequently submitted to US EPA for biennial reporting.

In South Carolina, for example, the state assigns profile numbers to generators. The facilities report quarterly to the State using the profile numbers and the State, in turn, reports biennially to EPA. South Carolina facilities do not have to complete a biennial Hazardous Waste Report, the State reports core data to EPA. In order for TSDFs to track waste streams to their original generators, the state profile number had to be used in conjunction with the EPA ID

Oregon also uses a profile number system rather than the EPA ID to track waste streams.

Texas uses a state waste code system and waste registration system that resulted in waste streams being aggregated in such a way that facilities had trouble matching waste streams from the BRS data with facility data. A few Texas respondents "did not recognize the" management or system types assigned by the State to the waste streams. Some Texas respondents determined that the waste stream for which information was requested was actually a combination of two or more waste streams. In several cases, the respondent provided constituent information on the individual waste streams comprising the waste stream reported in the Biennial Reporting System.

In general, respondents from states which did not use the 1993 Hazardous Waste Report for Biennial reporting may have been unable to identifying the waste streams in the survey or been unable to match the waste streams in the survey to the corresponding constituent data. Several respondents called the survey helpline with questions. Where possible, Westat faxed additional information to help the respondents identify the waste stream. Although state differences may have contributed to a facility not responding or not providing information for some waste streams, these differences did not in general prevent facilities from those states from responding.

10b. Confidential Business Information (CBI)

Three large facilities included in the NHWCS had filed their 1993 Hazardous Waste Reports as CBI. EPA IDs included in the CBI reports had been masked by EPA to protect confidentiality in order to include them in the National Oversight Database. Waste streams and EPA IDs selected for inclusion in the NHWCS also included these masked IDs and respondents had the additional

task of "unmasking" the ID to track the pertinent waste stream.

10c. Adding Constituents

For each waste stream, EPA provided a list of chemical constituents that were likely to be found in that waste stream. As noted above, respondents were provided with a blank page and a set of instructions and asked to add constituents that were not on the list. Approximately 70% of respondents did add constituents to at least one waste stream. Of those respondents who did add constituents, on average, additional constituents were added to about half of the waste streams.

The extent to which additional constituents could have been provided and were not is not known. There are documented calls to the help line in which respondents asked if they were to list all constituents or just those that were hazardous or toxic. Some of the added constituents were not hazardous. Also, the concentration data may not have been available for other constituents even through the respondents were aware that the constituent was present. Incinerators, for instance, may only have concentration measurements for metals and not for other constituents.

10d. Fuel blenders

Fuel blenders have a unique problem in that the survey requested information on characteristics and chemical constituents before the wastes were mixed for burning. However, fuel blenders typically measure the constituent concentrations in the mixture to be burned. Several fuel blenders assumed that constituents present in the mixture were present in all waste streams comprising the mixture if the constituent was listed as possibly present on the questionnaire.

10e. Handwriting Interpretation

Data are always subject to the interpretation of handwriting. Even with double key entry and verification, misinterpretations can occur.

10f. Batch Testing

The constituent concentrations may vary considerably over time within one waste stream, particularly for waste streams that consist of distinct batches of material. To the extent that any testing is performed before the waste is managed or treated, it is done on a batch basis to pre-qualify wastes. Characteristics and concentrations could vary enormously depending upon the batch. The survey questionnaire did not request the information needed to determine which waste streams have constituent concentrations which are relatively constant versus quite variable.

10g. Definition of "Derived From" Wastes

Question B14 required the respondent to estimate the percentage of the volume of waste that could be considered "derived from" hazardous waste. The intent was to use this percentage to estimate the quantity of hazardous waste derived from other hazardous waste through treatment, disposal or storage. A "derived-from" hazardous waste was defined in the Survey as any solid waste generated from the treatment, storage, or disposal of a hazardous waste, including any sludge, spill residue, ash, emission control dust, or leachate (but not including precipitation run-off). (40 CFR 261.3(c)(2)(i)).

In general, respondents had difficulty understanding this question and did not answer it consistently. This is because there are two different interpretations of the derived-from rule.

The first is a more literal, broad, and inclusive interpretation. The second is a very narrow interpretation. Then there are a number of other interpretations which would fall in between these broad and narrow interpretations.

The inclusive interpretation considers all wastes that come from a waste management system including blending to be a derived-from waste. Thus, all hazardous wastewater treatment sludges, leachates, and treatment residues would be considered derived-from waste. This would include listed wastes such as F039 (multi-source leachate). Listed wastewater treatment sludges (e.g. F006, K001) would also be considered derived-from wastes, even though they carry their own listing code. Waste that comes from fuel blending would also be considered derived-from wastes.

The narrow interpretation would consider a waste to be a derived-from waste only if it met the regulatory definition of a derived-from waste and would **not** be considered a hazardous waste on its own merits. If a waste stream met the listing definition of a hazardous waste (i.e., it carries its own listing code), it would not be considered derived-from waste. Thus F039 multi-source leachate would not be considered derived-from waste (although any waste generated from the treatment of the F039 leachate would be considered derived-from waste). Listed wastewater treatment sludges (e.g., F006, K001) also would not be considered derived-from wastes. Following this logic, then, characteristic wastes would not be considered derived-only wastes since they are considered hazardous wastes because of their chemical or physical properties. In addition, most mixtures (blending) of hazardous waste would not be considered derived-from wastes. In general, the only wastes that would be considered derived-from wastes are wastes (mostly treatment residuals) that carry a listed waste code solely due to 40 CFR 261.3(c)(2)(i).

During the QA/QC of the data, the broad interpretation of the derived-from rule was followed. Using the derived-from variable, the **secondary** waste generation can be segregated from the **primary** waste generation and each type of waste can then be studied separately if necessary.

11. Perspectives on Extrapolating the Data to Other Facilities

Users of the data may be interested in making conclusions about waste streams and facilities not responding to the survey or not covered by the survey. Regions within Figure 1 illustrate different combinations of waste stream types and facility types that may be of interest. The surface area of the regions in the figure are not proportional to the number of facilities or waste streams in a region or to the volume of waste represented by the facilities or waste streams within a region. The gray areas represent the data in the completed survey questionnaires. The arrows represent different extrapolations or scaling that might be performed. Extrapolation refers to making inferences based on assumptions. Scaling refers to making inferences justified by the sampling design.

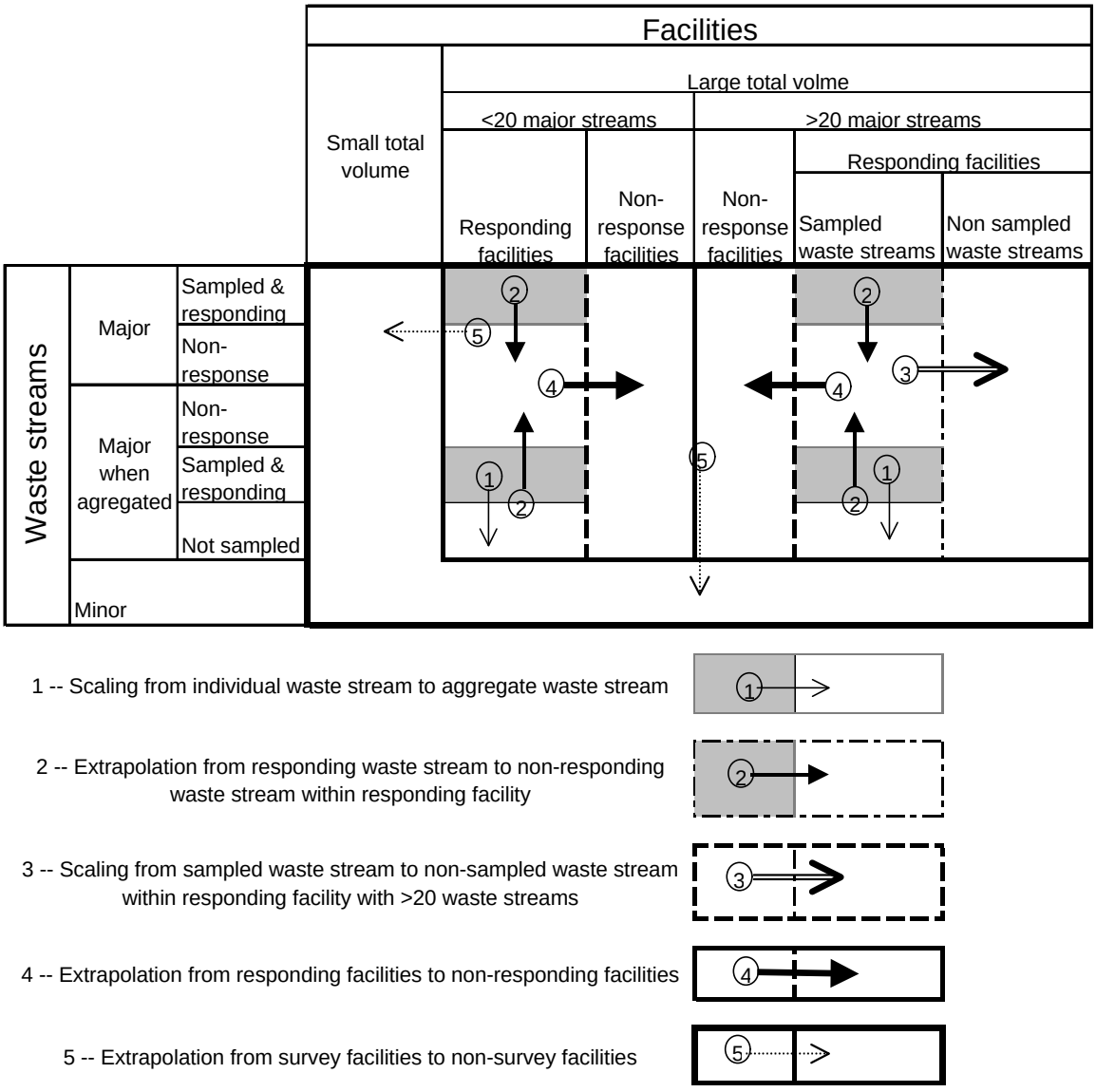
There are many different extrapolations or scaling that might be performed using this data. For example: 1 – Extrapolation from individual waste streams to aggregate waste streams, 2 – Extrapolation from responding waste streams to non-responding waste streams within a responding facility, 3 – Scaling from sampled waste streams to non-sampled waste streams within a responding facility with more than 20 major waste streams, 4 – Extrapolation from responding facilities to non-responding survey facilities, and 5 – Extrapolating from survey facilities to non-survey facilities.

1 – Extrapolation from individual waste streams to aggregate waste streams. For aggregate waste streams, one individual waste stream was randomly selected for the survey. AGGWGT

can be used to scale the volume of the individual waste stream up to the total for the aggregate waste stream. This scaling is shown in Figure 1 as a small arrow going from the sampled and responding waste streams to the non-sampled waste streams. This scaling is based on statistical sampling principles and should be done for most uses of the data. Scaling provides volume estimates for all waste streams (including aggregated streams) with completed questionnaires.

2 – Extrapolation from responding waste streams to non-responding waste streams within a responding facility. In some cases, the respondents provided information about some but not all of the waste streams in the survey. They may not have provided information on a waste stream because (1) they could not find the relevant information or (2) they believe the waste stream is not relevant to the HWIR process. In the first case, the data user might be willing to assume that the probability of locating the waste stream data is independent of the relevant characteristics of the waste stream. If so, the user might assume that the data for the missing waste stream was similar to the waste streams for which data was provided. In the second case, the respondent has indicated that the reason the data were not provided is probably related to the importance of the data to EPA.

Figure 1: Different combinations of waste streams and facilities which maybe important and extrapolations which may be used to estimate missing data.



It may not be reasonable to assume that the characteristics of the reported waste streams are similar to those for the missing waste stream. Any data that is available to assess if the missing data is similar to or different from the available data should be used to evaluate how to extrapolate from the reported waste streams to the missing waste streams.

In either case above, assumptions from outside of the survey process are required to extrapolate from the responding to the non-responding waste streams within a facility. As a general rule, use any available information to select the waste streams of interest and to identify similar waste streams that can be assumed to have similar characteristics. It may be that the missing waste streams are not of interest and can be ignored. With no information to decide, there are two primary options, (1) assume the non-responding waste streams are relevant and scale the responding waste streams up to the facility waste stream total based on the volumes or (2) ignore the non-responding waste streams. This requires making assumptions which may not be correct, and which are likely to add uncertainty to the result. For the NHWCS data, many users may

decide to ignore the waste streams with no data in facilities that responded, and assume these waste streams are not relevant to their analyses.

3 – Scaling from sampled waste streams to non-sampled waste streams within a responding facility with more than 20 major waste streams. Once estimates for all waste streams in the survey are obtained, these values can be scaled up to include the non-sampled waste streams in facilities that had more than 20 major waste streams. Since the sampling of waste streams was random, SAMPWGT can be used to scale the sampled waste streams up to get total or summary statistics for the facility. In Figure 1, this scaling is represented by an arrow with two lines and an open-style arrowhead. Note that extrapolation based on a random selection is statistically based.

4 – Extrapolation from responding facilities to non-responding survey facilities. Extrapolating from the responding survey facilities to the non-responding survey facilities (shown as the heaviest arrow crossing a heavy dashed line in Figure 1) also requires careful evaluation. Are there reasons to think that the responding facilities are different than the non-responding facilities, other than one responded and the other did not? The preloaded data may be used to determine if there are differences between the responding and non-responding facilities. Judgement must then be used to decide if those differences might affect differences in waste stream characteristics and constituents. Characteristics of the facilities or waste streams might be used to identify similar waste streams (or facilities) among the responding and non-responding facilities. Then the data from those waste streams in the responding facilities might be extrapolated to the similar waste streams in the non-responding facilities.

Because different users of the data are likely to want to extrapolate for different subsets of the data and a reasonable extrapolation for some purposes may not be the reasonable for others, EPA has not developed recommendations for extrapolating the survey responses to non-responding facilities. For example, a reasonable extrapolation for incinerator data may involve different factors than for wastewaters. Other users may be interested in other subsets of the data. Few users are likely to be interested in extrapolating to all facilities in the survey.

5 – Extrapolating from survey facilities to non-survey facilities. Finally, users may be interested in extrapolating to facilities not in the survey, for example small volume waste streams and facilities with small total volumes. Because these facilities are not included in the survey, the survey has no relevant information in a statistical sense.

Scaling from sampled waste streams to all waste streams based on random selection is supported by statistical theory. When scaling from the respondents to the non-respondents there is no formal statistical method of extrapolating. However, there is some reason to think that the respondents and non-respondents may be similar (at a minimum, they are similar in that they are all large volume facilities with major waste streams). In this case, assuming that the probability of responding is independent of the response may be reasonable. There is no statistical basis for extrapolation when scaling from the survey facilities to other facilities that would never have been in the survey. Extrapolation must be based on other information and judgement.

For example, suppose the user wants to extrapolate to waste streams with smaller volumes than those in the survey. Might the big volume waste streams in the survey be similar to the small volume waste streams of interest? Might the small individual waste streams that make up the aggregate waste streams be similar to the small waste streams of interest? Would concentrations in the small waste streams be higher or lower than in the large waste streams? Would the constituents be different? Does the 1986 generator survey have data that shows whether the

concentration in similar waste streams increases or decreases as the volume increases? Does a similar relationship exist between the large and much larger waste streams in the survey data? Is there reason to believe that the relationship across waste streams in the survey data also applies to smaller waste streams? Are there other data that might be used to develop a relationship between facilities similar to the survey facilities and facilities or waste streams of interest? What factors are likely to be related to the differences: volumes, waste characteristics, waste codes, management methods?

When extrapolating outside the range of the survey data, either additional information is required or assumptions must be made which need to be carefully considered. The results need to be qualified by presenting the assumptions. Random sampling adds uncertainty to an estimate, however, the amount of uncertainty can be estimated. The assumptions required for extrapolation also add additional uncertainty; and the magnitude of the uncertainty can be very large and difficult to assess.

Other extrapolation issues. Finally, when estimating an average or total volume, random variation (or uncertainty) in the data tends to add uncertainty but not bias to the estimates. For other uses, waste streams or facilities might be classified based on continuous measurements (such as concentrations). For example, waste streams may be classified as having concentrations of selected constituents greater than a selected cut-off. When the measurement has variation (sampling variation over time, measurement variation in the facility lab, perhaps sampling variation due to random selection of waste streams) the estimated percentage of facilities classified as over the cut-off will be biased. As a general rule, values at the extremes will be over-stated, that is, the estimates will be more extreme (high or low) than would be true with better data. If only 1% of facilities have values over the cut-off, the estimated percentage from the data might be 5% or 10% (depending on the magnitude of the errors).

12. CODING OF NUMERICAL VALUES

The data are entered into the data files to both reflect the answer provided by the respondent, as best as possible, and to create a data file which has consistent codes for each variable - codes which can be understood by a data user. In order to achieve these objectives, the following items were defined for each variable:

- A list of valid codes or a description of valid entries
- Rules for converting questionnaire responses to a common set of codes
- Quality control procedures to check the validity and reasonableness of the entries.

Respondent that exceeded the length of the variable were manually reviewed and assigned one of the valid codes.

Equivalent units were converted to a common unit of measure (for example, g/cc was converted to g/ml). When the respondent provided information in a form or format that was unexpected or inconsistent with the instructions, the respondent's entry was reviewed and coded to one of the common set of codes for the variable. The context provided by other responses was sometimes used to understand the unexpected entry. In some cases the respondent was contacted by phone for more information. For example, if the unit of measure for one constituent was not provided, concentrations for all other constituents had units of percent, and the concentration appeared to be reasonable when percent units were assumed, the units were assumed to be percent. If

multiple units were missing or different units were used, the respondent was called.

Constituent concentration information was considered to be most important. When the quality control procedures identified inconsistencies and problems in the constituent concentration data, the respondent was contacted to resolve the inconsistency or problem. During the call, inconsistencies and problems in other variables were also addressed.

The constituent concentrations were checked to verify that the total of all constituent concentrations (when expressed as percent) was less than 110%. In some cases the total exceeded 100% by a small amount which could be attributed to rounding. For other waste streams where the total exceeded 100%, the respondent was contacted to correct the problem. For many waste streams the constituent concentrations vary over time. Some respondents said that the reported concentrations represented upper bounds. The upper bound concentrations may total greater than 100%. For several cases where the total exceeded 110% and the respondent could not be reached, the concentrations were scaled, based on the characteristic information, so that the total was 100%. The variables MODLPPM and MODWPPM contain the ratios by which the reported whole waste and leachate concentrations, respectively, were multiplied to obtain a total of 100%.

In some cases, respondents reported a constituent concentration more than once in the same waste stream. EPA reviewed these cases and developed rules for resolving the duplicates. For the cases in the files, the rules corresponded to taking the maximum concentration across the duplicate records. The original data provided by the respondents are still in the data files. Only the summary concentrations and quantities (CWAVGVAL, CLAVGVAL, and CHEMMAS) were revised by zeroing values other than the maximum. The variable FFIXDUP indicates which records were affected.

Unit of measure codes were provided for quantity or volume information. In a few cases, the respondents used the volume units of measure instead of providing constituent concentrations. Thus they might report 234 tons per year of benzene. These values were converted to concentrations by dividing the quantities of constituents by the quantities of waste handled. In each case, the respondent was contacted to verify that this procedure would provide a reasonable concentration estimate.

The format for recording numbers in all variables except PCTDRVFR (percent of waste consists of "derived from" hazardous waste) and PCTNOHAZ (percent of the waste stream which is non-hazardous waste) are described below. The format for these variables is described in the list of variables.

Data format for concentrations and characteristic measurements

The following pieces of information were requested for numerical measures of constituent concentrations and physical characteristics (questionnaire tables B16 and C1):

- The **numerical value**, as either an average or central value, a range within which the average is expected to lie, an upper estimate for the average (such as a detection limit), NA (not applicable or not present), PR (present, not tested and concentration unknown), DK (don't know if the constituent is present) or ND (Not detected).

- The **units** of the numerical value. In some cases the units (percent by weight "%") were specified on the questionnaire. In most cases the respondent entered the units of measure.

- The **basis** for the numerical value. Numerical values based on fewer than 5 measurements or on estimates based on familiarity with the waste stream or engineering judgment were to be indicated by an "E". Values based on 5 or more measurements were to be indicated by an "A". The data were assumed to be from 1993. If data from other years was used to estimate the concentration of characteristics for the 1993 waste stream, the year from which the data were obtained was to be provided.

These numerical values were entered into the following four variables, where "XXX" stands for the type of measurement:

XXXA is the prefix to the number. Valid entries are:

<	Less than XXXB. XXXB may represent the upper end of a range, with zero or "not detected" as the low end of the range or XXXB may be the detection limit.
>	Greater than XXXB. Several responses provided a lower limit without an upper limit. For percentages, a reasonable upper limit is 100%.
NT	Not tested, but possibly present. A response that the concentration was "Unknown" was coded as NT. Respondents were contacted to determine if NT constituent concentrations indicated presence (PR) or absence (NA) of the constituent. Some respondents could not be reached.
PR	The constituent is believed by the respondent to be present in the waste at levels which might be of regulatory concern, however, the constituent concentration was not tested.
ND	Not detected, without a specified detection limit.
NA	Not applicable or not present. This code indicates that there is no reason to suspect the constituent is present. A concentration of "None" was coded as NA.
TR	Trace, greater than the detection limit, but a small enough value that numerical estimates are uncertain.
DK	The respondent does not know if the constituent is present in the waste stream or not.

XXXB the average, lower end of a range (when a value is in XXXD), or upper end of a range where the lower end is assumed to be zero, such as the detection limit.

XXXC a character to separate the endpoints of a range. The only valid entry in this field is "-" and only if there is a number in XXXD.

XXXD the upper end of a range.

The units of measurement are provided in the associated unit field, indicated here by the name XXXUNIT. XXXUNIT has units corresponding to the requested information. The units have been coded to have consistent entries for the same basic units. For example, since MG/KG and PPM are the same unit, MG/KG was recoded to PPM in the data files. For some characteristics, the units of percent by weight were specified on the questionnaire. For these characteristics, there is no variable for the unit of measure - percent by weight is assumed. In some cases the respondent crossed out the "%" and entered other units, such as PPM. In these cases the numerical value in XXXB was converted to percent. Specific gravity and pH have standard units. For pH and measures of "specific gravity or density", many respondents entered no units, NA, or S.U. (or some indication that standard units were used). If the measurement (XXXB) was reasonable for the standard unit of measure, the standard unit of measure was assumed.

Valid entries for the basis of measurement included an "A" or "E" and an indicator of the year from which data were available, such as "93", or "96". The order of these items and the punctuation might vary. When two years were specified, both are entered, such as "E9396", to indicate that the data were estimated from 1993 and 1996 data. If a range of years were specified a dash was used, for example, analysis of data from 1993 through 1996 would be coded as: "A93-6".

Using the symbol # to indicate a number, "UNIT" to indicate the presence of a valid unit of measure, the following patterns of entries in XXXA, XXXB, XXXC, XXXD, and XXXUNIT are considered valid:

XXXA	XXXB	XXXC	XXXD	XXXUNIT	Interpretation
NA				*	None, Not applicable
NT				*	Not tested, respondents were asked to recode "NT" constituent concentrations to PR, NA, or, if unable to do so, to DK
PR				*	Present at an unknown concentration
DK				*	The respondent does not know if the constituent is present in the waste stream
ND				*	Not detected, less than the detection limit, negative test
TR				*	Trace, greater than the detection limit, but a small enough value that numerical estimates are uncertain.
0				*	Zero, none, respondents were asked to recode these responses to PR, NA, or, if unable to do so, DK
<	#			UNIT	Less than #, # may be the detection limit or an upper value
<	#	-	##	UNIT	Average is in the range from non-detect (< #) to ##
	#			UNIT	# is an estimate of the average concentration
	#	-	##	UNIT	The average is between # and ##
>	#			UNIT	The average is greater than #

* A unit of measure may have been provided by the respondent.

For numerical characteristics and concentrations, the quality checks verified that the units were valid and that the entries in XXXA through XXXD were valid according to the table above. Additional quality control checks for individual variables are described below.

The constituent concentrations were converted to numerical values using the following assumptions:

- For ranges, use the midpoint of the range. For detection limits, use half the reported limit (the midpoint of zero and the reported value).
- For greater than values use the lower limit.
- For zero, missing, NA, NT, PR, ND, TR, DK, assume zero.

The numerical values are stored in variables WAVGVAL and CAVGVAL for the whole waste and leachate concentrations respectively. These values were converted to PPM, using the density estimate, in variables CWAVGVAL and CLAVGVAL. The concentrations were converted to chemical mass (tons) in the variable CHEMMASS. Adjustments to scale the totals constituent concentration to 100% or remove duplicate constituents affected only the variables CWAVGVAL, CLAVGVAL, and CHEMMASS.

13. DATA FILES

The data are in four text (ASCII) files. The same data are provided in four SAS files. One file (FACILTYx) contains information about the facilities for which respondent information is available. The second file (WSTREAMx) contains information about the waste streams (characteristic information, treatment system type information, and quantities handled). The third file (CONSTITx) contains constituent concentrations. The last file (COMMENTx) contains the comments. These files can be linked by the facility ID (EPAID), the waste stream number (WSNUM), and the constituent number (CONSTNUM). The letter "x" in the file names indicates a number which distinguishes different versions of the file. The latest files are FACILTY5, WSTREAM5, CONSTIT5, and COMMENT5.

Appendix 3 and 4 describe the contents of the variables in the data files. Appendix 3 lists the variables in the order they occur in the questionnaire with a detailed explanation of how each variable was coded. Appendix 4 contains an alphabetic list of the variables in each file with a brief description of their contents.

The facility, waste stream, and constituent files have two types of information, the preloaded information provided on the questionnaire and the data provided by the respondents. If the respondents did not change the preloaded information, the variables with the respondent information show the preloaded data. The variables containing the preloaded data have variable names that end in X. By comparing the preloaded and respondent data, the user can determine which preloaded data items were changed by the respondent.

14. APPENDICES

Appendix 1 contains the facilities that responded to the Survey but their data were not verified.
Appendix 2 contains the item-by-item response rate.
Appendix 3 contains a list of the variables in the data files.
Appendix 4 contains the file format for the data files.
Appendix 5 contains the documentation for the Rollins Corporation files.

Appendix 1

List of sites that responded to the NHWCS but their data were not verified.

No.	EPA ID	Site Name	City
1.	ILD010284248	CID Recycling and Disposal Facility	Calumet City
2.	MID005358130	Total Petroleum, Inc.	Alma
3.	MID096963194	Chem-Met Services Inc.	Wyandotte
4.	MOD981127319	Lone Star Industries, Inc.	Cape Girardeau
5.	OHD005108477	Aristech Chemical Corp.	Haverhill
6.	PAD002334753	Occidental Chemical Corp.	Pottstown

7. PRD090399718 Safety Kleen Manati

8. SC1890008989 DOE/Savannah River Aiken

9. SCD070375985 Laidlaw Pinewood

10. TN3890090001 U.S. DOE Y-12 Plant Oak Ridge

11. WVD004325353 OSI Specialties, Inc. Sisterville

These sites were contacted repeatedly to discuss their submissions but the cases remained unresolved at the time the QA/QC process was stopped.

Appendix 2

Appendix 3

Variables, valid entries, and QC checks

For each variable, the following outline describes the source of the information, the valid entries, how the entries were recoded, and any QC checks. After presenting the variables which link the files and characterize the facilities and waste streams, the variables are presented in the order in which the data is requested in the questionnaire. The data files contain both the preloaded data from the BRS system and the final data, as provided or modified by the respondent. The variable names for the preloaded information are obtained by changing the last character of the variable name to an X.

The variables which link the files are the EPA ID, the waste stream number (WSNUM) within each facility, and the constituent number within each waste stream (CONSTNUM).

EPA ID No.

EPAID_A4

Source: Obtained from the BRS system.
Valid Entries: Entries selected for the survey were considered valid.
QC checks: All EPA IDs are valid..

Waste Stream Number

WSNUM

Source: Created when the data were entered. This number is the order in which the waste streams were listed in the survey questionnaire. In cases where the respondent provided information on more waste streams than requested in the survey, the additional waste streams were appended to the end of the survey and numbered sequentially from the last requested waste stream. In particular, where the respondent provided information on two waste streams which were aggregated by the State of Texas, the first waste stream is associated with the requested waste stream, the second is appended to the end of the survey (see WSTYPE). The order has no significance. The waste stream number is used to link the constituent concentrations to the characteristic and volume information for each waste stream.
Valid Entries: numbers between 1 and 20. For records which apply to facilities (in FACILTYx and COMMENTx, WSNUM = 0.
QC checks: All waste stream numbers are valid.

Constituent Number

CONSTNUM

Source: Created when the data were entered. This number is the order in which the constituents were listed in the questionnaire (either preloaded or entered by the respondent). The order has little significance, for preloaded constituents, the constituents were listed alphabetically by constituent name.
Valid Entries: Consecutive integers starting at 1 up to the number of constituents listed for the waste stream. For records FACILTYx, WSTREAMx, and records in COMMENTx which refer to the facility or waste stream, CONSTNUM = 0.
QC checks: All constituent numbers are valid.

The data files contain three variables derived from the BRS which can be used to link to the BRS data files and also merge the facility and waste stream data. Within each EPAID, the order of the waste streams in the survey corresponds to sorting by GM_WR BRS_PAGE and BRS_SUBP.

BRS page

BRS_PAGE

Source: BRS.
Valid Entries: Integers
QC checks: Checked for missing values. The BRS_PAGE and BRS_SUBP are missing for waste streams which were added by the respondents.

BRS subpage

BRS_SUBP

Source: BRS.
Valid Entries: Integers
QC checks: Checked for missing values. The BRS_PAGE and BRS_SUBP are

missing for waste streams which were added by the respondents.

BRS form

GM_WR

Source: BRS.

Valid Entries: Either GM (for waste generated and managed) or WR for waste received from off-site.

QC checks: Checked for missing values. The BRS_PAGE and BRS_SUBP are missing for waste streams which were added by the respondents.

Four weight variables have been added.

BRS reporting flag

BRSWGT

Source: Respondents notes.

Valid Entries: Valid entries are:

0 Indicates the waste stream should not have been in the BRS

1 Indicates the respondent did not question that the waste stream was in the BRS data system

QC checks: No missing values.

Indicator of survey data versus volunteered data

INSURWGT

Source: Respondents notes.

Valid Entries: Valid entries are:

0 Indicates the waste stream data were not requested in the survey questionnaire

1 Indicates the waste stream data were requested in the survey questionnaire

QC checks: No missing values.

Sampling weight for scaling up from the sampled waste streams (including aggregated streams) to all waste streams at a facility.

SAMPWGT

Source: Sampling procedure programs and files

Valid Entries: Valid entries are:

A number greater than or equal to 1.

QC checks: No missing values.

Sampling weight for scaling up from the sampled individual waste streams to all waste streams with similar characteristics which make up an aggregated waste stream.

AGGWGT

Source: Respondents notes.

Valid Entries: Valid entries are:

0 Indicates the waste stream should not have been in the BRS

1 Indicates the respondent did not question that the waste stream was in the BRS data system

QC checks: No missing values.

Two variables have been added to help the users of the data. These variables provide information which qualifies the facility response and the waste streams data.

Facility type

FACTYPE

Source: Survey implementation records.

Valid Entries: Valid entries are:

Survey	Indicates the facility was sent a survey questionnaire
Pretest	Indicates the facility was sent a pretest survey questionnaire
Rollins	Indicates the data were provided by Rollins in a computer file based on the questions in the pretest survey.

QC checks: No missing values.

Waste stream type

WSTYPE

Source: Survey implementation records.

Valid Entries: Valid entries are:

Only Shipped	Indicates the waste was shipped and not managed by the facility (as indicated in the BRS)
Split..	Indicates the waste stream requested in the survey was considered to be two different waste streams by the respondent. The first of those waste streams has WSTYPE = Split nA, where n is the number of the waste stream requested in the survey. The second waste stream has WSTYPE = Split nB.
Fuel Blending	Indicates that the characteristic data are based on the waste as received and the constituent data are for the blended waste.
Proprietary	Indicates that the waste stream data was considered proprietary and was not provided.
Reporting error	Indicates that the waste stream was incorrectly reported to BRS.
Two streams	Indicates that the reported values are for two streams, the requested stream and a second stream. In this case, the requested stream made up the vast majority of the volume of the two streams reported on.
BRS dup	Indicates that the wastes stream was in the BRS twice and that both BRS waste streams were sampled for the survey.
Duplicate	Indicates that the waste stream was mistakenly duplicated in one pretest questionnaire.

QC checks: None.

Two variables have been added to help merge the comments with the data.

Question order

QORDER

Source: Survey comment processing.

Valid Entries: Integers

QC checks: All entries are valid.

Comment: This numeric field is used to order the comments in the order of the questions to which they refer. This may be useful for printing the comments.

Comment sequence

SEQ

Source: Survey comment processing.

Valid Entries: Integers

QC checks: All entries are valid.
Comment: For comments which exceed 200 characters and therefore are on multiple records, this numeric field is used to sort the comment records.

The following information was requested on the questionnaire.

Part I: General Facility Information Overview

A1. Facility Name
FACNAME
Source: Obtained from the BRS system unless modified by the respondent.
Valid Entries: All entries were considered valid.
QC checks: None.
Comments: preloaded variable is FACNAMEX. In some cases the facility name has changed from that listed on the 1993 BRS submittal.
Related variables: FACNAMEX

A2. Location Address
LADDRESS Street Address
LCITY City
LSTATE State
LZIPCODE Zip code
Source: Obtained from the BRS system unless modified by the respondent.
Valid Entries: All entries were considered valid.
QC checks: None. All addresses present however two zip codes are missing.
Comments: In one case the facility location changed from that on the 1993 BRS because the post office changed the town designation and zip code.
Related variables: LADDRESX, LCITYX, LSTATEX, LZIPCODX

A3. Mailing Address
MADDRESS Street Address
MCITY City
MSTATE State
MZIPCODE Zip code
Source: Obtained from the BRS system unless modified by the respondent.
Valid Entries: All entries were considered valid.
QC checks: None.
Related variables: MADDRESX, MCITYX, MSTATEX, MZIPCODX

A5. TRI No. TRI number
TRID
Source: Blank unless added by the respondent.
Valid Entries: All entries were considered valid.
QC checks: None.

A6. NPDES Permit No.
NPDES
Source: Blank unless added by the respondent.
Valid Entries: All entries were considered valid.
QC checks: None.

A7. Dun & Bradstreet No.

DNB

Source: Blank unless added by the respondent.
Valid Entries: All entries were considered valid.
QC checks: None.

A9. 1993 Hazardous Waste Report Respondent Information

RESPNDT Name
RESPTITL Title
RESPPHON Telephone

Source: Obtained from the BRS system unless modified by the respondent.
Valid Entries: All entries were considered valid.
QC checks: None.
Comments: In several cases the respondent changed this information, perhaps to indicate that the current contact for questions about the 1993 BRS data has changed.
Related variables: RESPNDTX, RESPTITX, RESPPHOX

A9. National Hazardous Waste Constituent Survey Respondent Information

CNSTUNT Name
CNSTTITL Title
CNSTPHON Telephone

Source: Obtained from the respondent.
Valid Entries: All entries were considered valid.
QC checks: None.
Comments: In six cases the respondent information is missing. In three cases, no respondent information was provided. In three cases the respondent was apparently the same as the contact for the 1993 BRS report (see above).

Part II Waste Stream-Specific Information

[Page 2 top]

WST_ORG Waste Origin

Source: Obtained from the BRS system unless modified by the respondent.
Valid Entries:
Generated on Site
Received From XX [where XX is a two letter state abbreviation]
QC checks: The values were checked to against the list of valid codes, all entries were valid.
Related variable: WST_ORGX

WSTCODES Waste Code(s) [EPA Hazardous waste codes]

Source: Obtained from the BRS system unless modified by the respondent.
Valid Entries: EPA waste codes from Attachment 7 to the survey instructions.
QC checks: None.
Comments: Although this variable is an important identifier of the waste stream, the respondents may have changed the EPA waste codes. The changes are assumed to reflect a more accurate description of the waste.
Related variables: WSTCODEX, FWSTCODC, FWSTCODF

STATE_CD State waste codes

Source: Obtained from the BRS system unless modified by the respondent.
Valid Entries: All entries were considered valid.

QC checks: None.
Related variable: STATE_CDX

WSTDESC Waste Description

Source: Obtained from the BRS system unless modified by the respondent.
Valid Entries: All entries were considered valid.
QC checks: None.
Related variable: WSTDESCX

B. Waste Quantities and Characteristics

B1. Waste Form Code

FORMCODE Waste Description

Source: Obtained from the BRS system unless modified by the respondent.
Valid Entries: Valid entries are EPA form codes in Attachment 5 to the survey instructions and:

DK	Do not know
N/A	Not Applicable
missing	

QC checks: The values were checked to against the list of valid codes. Invalid entries that were taken from the BRS system and were not updated by the respondent remain in the data files. All the invalid codes come from Texas.
Related variables: FORMCODX

B2. Is this waste stream a wastewater?

WW

Source: This field was calculated based on BRS information. The respondent may have corrected the field.
Valid Entries:

Y	Yes
N	No

QC checks: Consistency with other fields
Related variables: WWX

B3. Was this stream managed exclusively in units exempt from RCRA permitting requirements?

RCRAEXMP

Valid Entries:

Y	Yes
N	No

Source: This field was determined from BRS information. The respondent may have corrected the field. This question was not asked of the pretest facilities. Therefore, this field will be blank for pretest facilities.
QC checks: All entries were valid.
Related variable: RCRAEXMX

B4. System types in which the waste stream was managed. The questionnaire allowed for entering up to 12 system types. No more than 4 system types use reported for any waste stream. In the data file, variables are provided for up to 5 system types.

SYSTDRn System [n = 1 to 5]

Valid Entries: System type codes from Attachment 6 of the survey instructions.

SYSQTYn Quantity Managed [n = 1 to 5]

Valid Entries: Numbers

UNITSn Unit of measure [n = 1 to 5]

Unit of measure codes from an extra page included with the survey instructions,

1	Pounds
2	Short tons (2000 pounds)
3	Kilograms
4	Metric tonnes (1,000 kilograms)
5	Gallons
6	Liters
7	Cubic yards
N/A	Not Applicable

Source: This field was determined from BRS information. The respondent may have corrected the field.

Related variables: FSYSTDR1, FSYTDR1X, FWSTCODX, FSYSQTYX, CSYQTYnX, CSYSQTY, CSYSQTYn, CSYSQTYX, SYSQTY, SYSQTYnX, SYSQTYX, SYSTDRnX, UNIT, UNITnX, UNITX

CSYSQTY is the numerical quantity managed, in tons.

B5. SIC Code - For waste generated on site only

SICCODE

Source: Obtained from the BRS system unless modified by the respondent.

Valid Entries: System type codes from Attachment 2 of the survey instructions and "N/A".

QC checks: None.

Related variables: SIC, RCFNSIC, SICCODEX, NHWCSSIC,

SIC is the SIC code of the generator of the waste.

B6. Source Code - For waste generated on site only

SRCCODE

Source: Obtained from the BRS system unless modified by the respondent.

Valid Entries: Source codes from Attachment 4 of the survey instructions and:

N/A	Not Applicable

QC checks: The values were checked to against the list of valid codes. Invalid entries that were taken from the BRS system ("A") and were not updated by the respondent remain in the data files. One invalid code (B33) were entered by the respondent.

Related variable: SRCCODE

B7. Origin Code - For waste generated on site only

ORIGCODE

Source: Obtained from the BRS system unless modified by the respondent.

Valid Entries: Origin codes from Attachment 3 of the survey instructions and:

6	Undefined code, preloaded in pretest questionnaires only
DK	Do not know
N/A	Not Applicable

QC checks: The values were checked to against the list of valid codes.

Related variable: SRCCODEX

B8. Point of measurement Code - For waste generated on site only

POMCODE

Source: Obtained from the BRS system unless modified by the respondent.

Valid Entries: Point of measure codes from Attachment 3 of the survey instructions and:

DK Do not know
N/A Not Applicable

QC checks: The values were checked to against the list of valid codes.

Comments: In some cases the point of measure codes for the quantity, characteristics, and constituents are different.

Related variable: POMCODEX

B9. Quantity of Waste Generated - For waste generated on site only

QUAN1993

Source: Obtained from the BRS system unless modified by the respondent.

Valid Entries:

Numbers greater than or equal to zero.

-1.00 indicates that the waste stream was received from off-site according to the BRS data

QC checks: Values of zero and invalid entries were checked. Two zero quantities were preloaded.

Related variable: QUAN199X

B10. Unit of Measure - For waste generated on site only

QGENUNIT

Source: Obtained from the BRS system unless modified by the respondent.

Valid Entries:

Unit of measure codes from an extra page included with the survey instructions,

1 Pounds
2 Short tons (2000 pounds)
3 Kilograms
4 Metric tonnes (1,000 kilograms)
5 Gallons
6 Liters
7 Cubic yards
N/A Not Applicable

QC checks: The values were checked to against the list of valid codes and that valid quantities (QUAN1993) have units of measure.

Related variable: QGENUNIX

B11. Off-site Facility's EPA ID Number - For waste received from off--site only

OFFEPAID

Source: Obtained from the BRS system unless modified by the respondent.

Valid Entries: All entries beginning with a two letter state code, "ID" (IDGEN or IDCBI indicate EPA ID's considered to be CBI), or "FC" (for Foreign Country) with a total of 12 characters were considered valid.

QC checks: Missing entries were replaced with valid entries where possible.

Related variable: OFFEPAIX

B12. Quantity of Waste Generated - For waste received from off--site only

QUANRECV

Source: Obtained from the BRS system unless modified by the respondent.

Valid Entries:

Numbers greater than or equal to zero.

-1.00 indicates that the waste stream was received from off-site according to the BRS data

QC checks: None.
Related variable: QUANRECX

B13. Unit of Measure - For waste received from off--site only
QRECUNIT
Source: Obtained from the BRS system unless modified by the respondent.
Valid Entries:
Unit of measure codes from an extra page included with the survey instructions,

1	Pounds
2	Short tons (2000 pounds)
3	Kilograms
4	Metric tonnes (1,000 kilograms)
5	Gallons
6	Liters
7	Cubic yards
N/A	Not Applicable
NA	Not Applicable (entered by the respondent)

QC checks: The values were checked to against the list of valid codes and that valid quantities (QUANRECV) have units of measure.
Related variable: QRECUNIX

B14. What percentage of this waste stream consists of "derived from" hazardous waste?
PCTDRVFR
Source: Survey respondent.
Valid Entries:
Numbers or ranges between 0 and 100

NA	Not applicable or not available
NT	Not tested, Unknown, or Don't Know

QC checks: The questionnaire entries were checked to against the list of valid codes and recoded where possible.
Recode DK, UNKNOWN to NT.

B15. What percentage of this waste stream consists of non-hazardous waste?
PCTNOHAZ
Source: Survey respondent.
Valid Entries:
Numbers or ranges between 0 and 100

NA	Not applicable or not available
NT	Not tested, Unknown, or Don't Know

QC checks: The questionnaire entries were checked to against the list of valid codes and recoded where possible.
Recode DK, UNKNOWN to NT.

B16. What is the physical and chemical form of the waste stream?
1. Specific gravity or density
Source: Survey respondent.
SPECGRA[A,B,C,D]
Valid Entries: Non-negative numbers as one value or a range, NA, ND, NT, or TR. Responses of "Variable" for two waste streams were coded to NT (equivalent to unknown).
QC checks: Number were checked for reasonableness. Reasonable values were

assumed to be within a factor of three of the density of water for the reported units.

SPGRUNIT

Valid Entries:

#/FT3	Pounds per cubic foot
#/GAL	Pounds per gallon
#/Y3	Pounds per cubic yard
G/ML	grams per milliliter, grams per cubic centimeter, G/CC, GM/CC, GM/ML
g/ml	No unit was provided and grams per milliliter was assumed after reviewing the data
G/L	Grams per liter
T/Y3	Tons per cubic yard
blank	If the specific gravity is zero

QC checks: All values were recoded to valid entries. An indication that the unit was specific gravity (e.g. SP.GR.) was recoded to g/ml. Because there are standard units for specific gravity, responses with no units provided and numerical entries near 1.0 for water and organic solvent mixtures were recoded to g/ml.

Recode SPGRUNIT from:

G/CC	to	G/ML
GM/CC	to	G/ML
GM/L	to	G/L
GM/ML	to	G/ML
S.G.	to	g/ml
SG	to	g/ml
SGH2O	to	g/ml
SP GR	to	g/ml
SP.G.	to	g/ml
SP.GR	to	g/ml
SU	to	g/ml

SPGRBASI

Valid Entries: Valid entries for basis of measurement

QC checks: All values were recoded to clearly indicate A (analysis) or E (estimate) and a year, if provided. Values were not recoded for consistency

Related variables: CBRSSPGR, BRSDENSX, SPGRNEW, SPECGRA, CSPECGRA, BRSDUOMX, SPGRNEW combines the specific gravity estmates from different sources.

2. Percent solids by weight

Source: Survey respondent.

PCTSOLI[A,B,C,D]

Valid Entries: Non-negative numbers between 0 and 100 as one value or a range, NA, ND, NT, or TR

QC checks: The questionnaire entries were checked to against the list of valid codes and recoded where possible.

PCTBASI

Valid Entries: Valid entries for basis of measurement

QC checks: All values were recoded to clearly indicate A (analysis) or E (estimate) and a year, if provided. Values were not recoded for consistency

3. Percent total suspended solids by weight

Source: Survey respondent.

PCTTSS[A,B,C,D]

Valid Entries: Non-negative numbers between 0 and 100 as one value or a range, NA, ND, NT, or TR

QC checks: The questionnaire entries were checked to against the list of valid codes and recoded where possible.

TSSBASIS

Valid Entries: Valid entries for basis of measurement

QC checks: All values were recoded to clearly indicate A (analysis) or E (estimate) and a year, if provided. Values were not recoded for consistency

4. Percent ash by weight

Source: Survey respondent.

PCTASH[A,B,C,D]

Valid Entries: Non-negative numbers between 0 and 100 as one value or a range, NA, ND, NT, or TR

QC checks: The questionnaire entries were checked to against the list of valid codes and recoded where possible.

ASHBASIS

Valid Entries: Valid entries for basis of measurement

QC checks: All values were recoded to clearly indicate A (analysis) or E (estimate) and a year, if provided. Values were not recoded for consistency

5. Percent water by weight

Source: Survey respondent.

PCTWATE[A,B,C,D]

Valid Entries: Non-negative numbers between 0 and 100 as one value or a range, NA, ND, NT, or TR

QC checks: The questionnaire entries were checked to against the list of valid codes and recoded where possible.

H2OBASIS

Valid Entries: Valid entries for basis of measurement

QC checks: All values were recoded to clearly indicate A (analysis) or E (estimate) and a year, if provided. Values were not recoded for consistency

6. Percent total organic carbon by weight

Source: Survey respondent.

PCTTOC[A,B,C,D]

Valid Entries: Non-negative numbers between 0 and 100 as one value or a range, NA, ND, NT, or TR

QC checks: The questionnaire entries were checked to against the list of valid codes and recoded where possible.

TOCBASIS

Valid Entries: Valid entries for basis of measurement

QC checks: All values were recoded to clearly indicate A (analysis) or E (estimate) and a year, if provided. Values were not recoded for consistency

7. Percent oil and grease by weight

Source: Survey respondent.

PCTOIL[A,B,C,D]

Valid Entries: Non-negative numbers between 0 and 100 as one value or a range,

NA, ND, NT, or TR

QC checks: The questionnaire entries were checked to against the list of valid codes and recoded where possible.

OILBASIS

Valid Entries: Valid entries for basis of measurement

QC checks: All values were recoded to clearly indicate A (analysis) or E (estimate) and a year, if provided. Values were not recoded for consistency

8. pH

Source: Survey respondent.

PH[A,B,C,D]

Valid Entries: Non-negative numbers or a range of numbers generally within the range of 0 to 14, NA, ND, NT, or TR

QC checks: Invalid entries were checked and recoded if possible based on a review of the questionnaire. All numerical values were in the range from 0 to 14.

PHUNIT

Valid Entries:

S.U.	Standard pH units

Recode all units to S.U.

QC checks: All values were recoded to valid entries. Because there are standard units for pH, responses with no units provided were recoded to S.U.

PHBASIS

Valid Entries: Valid entries for basis of measurement

QC checks: All values were recoded to clearly indicate A (analysis) or E (estimate) and a year, if provided. Values were not recoded for consistency

9. Flash point

Source: Survey respondent.

FLASH[A,B,C,D]

Valid Entries: Non-negative numbers or a range, NA, ND, NT, or TR. NEG (used by one respondent) apparently indicates that the waste does not burn.

QC checks: The questionnaire entries were checked to against the list of valid codes and recoded where possible.

FLSHUNIT

Valid Entries:

C	Centigrade
F	Fahrenheit

QC checks: All values were recoded to valid entries. Because there are standard units for pH, responses with no units provided were recoded to S.U.

Recode FLSHUNIT from:

DEG F	F
DEG.F	F
FAHR.	F

FLSHBASI

Valid Entries: Valid entries for basis of measurement

QC checks: All values were recoded to clearly indicate A (analysis) or E (estimate) and a year, if provided. Values were not recoded for consistency

10. BTU Content (measured using standard test)

Source: Survey respondent.

BTU[A,B,C,D]

Valid Entries: Non-negative numbers or a range, NA, ND, NT, or TR

QC checks: The questionnaire entries were checked to against the list of valid codes and recoded where possible.

BTUUNIT

Valid Entries:

/LB	BTU per pound
/GAL	BTU per gallon
/L	BTU per liter

QC checks: The questionnaire entries were checked to against the list of valid codes and recoded where possible.

Recode BTUUNIT from:

/LB.	to	/LB
BTU #	to	/LB
BTU/#	to	/LB
BTU/L	to	/L
GAL	to	/GAL
L	to	/L
LB	to	/LB
LBS	to	/LB
POUND	to	/LB

BTUBASIS

Valid Entries: Valid entries for basis of measurement

QC checks: All values were recoded to clearly indicate A (analysis) or E (estimate) and a year, if provided. Values were not recoded for consistency

11. Total organic halogen (percent by weight)

Source: Survey respondent.

HALOGEN[A,B,C,D]

Valid Entries: Non-negative numbers between 0 and 100 as one value or a range, NA, ND, NT, or TR

QC checks: The questionnaire entries were checked to against the list of valid codes and recoded where possible.

HALOBASI

Valid Entries: Valid entries for basis of measurement

QC checks: All values were recoded to clearly indicate A (analysis) or E (estimate) and a year, if provided. Values were not recoded for consistency

12. Five-day biological oxygen demand concentration

Source: Survey respondent.

BIO5DAY[A,B,C,D]

Valid Entries: Non-negative numbers or a range, NA, ND, NT, or TR

QC checks: The questionnaire entries were checked to against the list of valid codes and recoded where possible.

BIO5UNIT

Valid Entries:

MG/L	Milligrams per liter
PPM	Parts per million
#/DAY	Pounds per day

QC checks: The questionnaire entries were checked to against the list of valid codes and recoded where possible.

BIO5BASI

Valid Entries: Valid entries for basis of measurement
QC checks: All values were recoded to clearly indicate A (analysis) or E (estimate) and a year, if provided. Values were not recoded for consistency

13. Chemical oxygen demand concentration
Source: Survey respondent.
CEHMOX[A,B,C,D]
Valid Entries: Non-negative numbers or a range, NA, ND, NT, or TR
QC checks: The questionnaire entries were checked to against the list of valid codes and recoded where possible.
CHOXUNIT
Valid Entries:
MG/L Milligrams per liter
PPM Parts per million
QC checks: The questionnaire entries were checked to against the list of valid codes and recoded where possible.
CHOXBASI
Valid Entries: Valid entries for basis of measurement
QC checks: All values were recoded to clearly indicate A (analysis) or E (estimate) and a year, if provided. Values were not recoded for consistency

B15. Point of measurement code for point at which above characteristics were measured
PHYSCHAR
Source: Survey respondent.
Valid Entries: Point of measure codes from Attachment 3 of the survey instructions and "N/A".
QC checks: The questionnaire entries were checked to against the list of valid codes and recoded where possible.
Comments: In some cases the point of measure codes for the quantity, characteristics, and constituents are different.

C. Constituent Concentration Data

C1. Average concentration during 1993 on both whole waste and leachate basis
1. CAS number and constituent name
CAS
Valid Entries: Valid CAS numbers with dashes, or blank
QC checks: Checked for consistency with the constituent name and a list of valid CAS-constituent pairs. When in doubt, the respondent's CAS number was assumed to be correct.
CONSTIT
Valid Entries: Chemical names, with the preferred name first, followed by acceptable alternate names in square brackets.
QC checks: Checked for consistency with the CAS number and a list of valid CAS-constituent pairs. When in doubt, the respondent's CAS number was assumed to be correct.
Related variable: CASR, CONSTITR

2. Whole Waste Concentration
Source: Survey respondent.
WAVGVAL[A,B,C,D]

Valid Entries: Non-negative numbers or a range, NA, ND, NT, DK, PR, or TR

QC checks: Invalid entries were checked and recoded if possible based on a review of the questionnaire. The respondent was contacted to resolve discrepancies where the data could not be recoded to a valid entry.

Recode WAVGVALA from:

<##	to	ND	
N/A	to	NA	
NONE	to	NA	
NT<	to	<	[if there is a number in WAVGVAL]
ND-	to	ND	[if there is a no number in WAVGVAL]

Recode WAVGVALC from:

TO	to	-

WWUNIT

Valid Entries:

UG/L	Micrograms per liter
MG/L	Milligrams per liter
PPM	Parts per million, MG/KG
PPB	Parts per million, UG/KG
V%	Percent by volume, volume percent
%	Percent. percent by weight, weight percent

QC checks: Invalid entries were checked and recoded if possible based on a review of the questionnaire. The respondent was contacted to resolve discrepancies where the data could not be recoded to a valid entry.

Recode WWUNIT from:

% WT	to	%
%(W)	to	%
%(WT)	to	%
%V	to	V%
%WT	to	%
MG/KG	to	PPM
UG/KG	to	PPB
W%	to	%
WT %	to	%
WT%	to	%
WT. %	to	%
WT.%	to	%

WWBASIS

Valid Entries: Valid entries for basis of measurement

QC checks: All values were recoded to clearly indicate A (analysis) or E (estimate) and a year, if provided. Values were not recoded for consistency

Related variable: CWAVGVAL, WAVGVAL, CHEMMASS

3. Leachate Concentration

Source: Survey respondent.

LAVGVAL[A,B,C,D]

Source: Survey respondent.

Valid Entries: Non-negative numbers or a range, NA, ND, NT, DK, PR, or TR

QC checks: Invalid entries were checked and recoded if possible based on a review of the questionnaire. The respondent was contacted to resolve discrepancies where the data could not be recoded to a valid entry.

LCUNIT

Valid Entries:

UG/L	Micrograms per liter
MG/L	Milligrams per liter
PPM	Parts per million, MG/KG
PPB	Parts per million, UG/KG
V%	Percent by volume, volume percent
%	Percent. percent by weight, weight percent

QC checks: Invalid entries were checked and recoded if possible based on a review of the questionnaire. The respondent was contacted to resolve discrepancies where the data could not be recoded to a valid entry.

LCBASIS

Valid Entries: Valid entries for basis of measurement

QC checks: All values were recoded to clearly indicate A (analysis) or E (estimate) and a year, if provided. Values were not recoded for consistency

Related variable: CLAVGVAL, LAVGVAL, CHEMMASS

C2. Point of measurement code at which constituent concentrations were measured

C2

Source: Survey respondent.

Valid Entries: Point of measure codes from Attachment 3 of the survey instructions and "N/A".

QC checks: The values were checked to against the list of valid codes. The respondent was contacted to resolve discrepancies where the data could not be recoded to a valid entry.

Comments: In some cases the point of measure codes for the quantity, characteristics, and constituents are different. This variable is in the waste stream file.

The following variables are in the comment file. Explanatory text from the respondents were typed into the comments and associated with a question in the questionnaire. These comments will be reviewed and some of these comments will be used to revise the survey data.

Comment text

COMMENT

Source: Respondent.

Valid Entries: Any text. If the text exceeded 200 characters, multiple comment records were used.

QC checks: None.

Question number corresponding to the comment

QUES

Source: Survey processing.

Valid Entries: A question number or code indicating a reference for the comment: 1A, B1, B3, B4, B6, B10, B14, B15, B16, B17, C1, C2, WC (waste codes), WD (waste description), and WO (Waste origin)

QC checks: None.

Appendix 4

Variables listed alphabetically within each file. The suffix "x" on the file name distinguishes different versions of the files.

File name: FACILTYx

Variable name	Description
CNSTPHON	A9 NHWCS respondent's phone number
CNSTTITL	A9 NHWCS respondent's title
CNSTUNT	A9 NHWCS respondent's name
CONSTNUM	Constituent number used for ordering comments by constituent when printing comments (all values are 0 in the FACILTYx file)
DNB	A7 Dun & Bradstreet number
EPAID	EPAID
FACNAME	A1 Facility name
FACNAMEX	A1 Facility name (preloaded)
FACTYPE	Facility type (pretest or main survey)
HASCCONC	A flag (Y or N) indicating if the respondent provided positive constituent concentrations for at least one constituent in at least one waste stream which was managed on-site.
LADDRESS	A2 Facility location address
LADDRESX	A2 Facility location address (preloaded)
LCITY	A2 Facility location city
LCITYX	A2 Facility location city (preloaded)
LSTATE	A2 Facility location state
LSTATEX	A2 Facility location state (preloaded)
LZIPCODE	A2 Facility location zip code
LZIPCODX	A2 Facility location zip code (preloaded)
MADDRESS	A3 Facility mailing address
MADDRESX	A3 Facility mailing address (preloaded)
MCITY	A3 Facility mailing city
MCITYX	A3 Facility mailing city (preloaded)
MSTATE	A3 Facility mailing state
Variable name	Description
MSTATEX	A3 Facility mailing state (preloaded)
MZIPCODE	A3 Facility mailing zip code
MZIPCODX	A3 Facility mailing zip code (preloaded)
NPDES	A6 NPDES permit number
QORDER	Used for ordering comments by question within section when printing comments (all values are 0 in the FACILTYx file)
RESPNDT	A8 1993 Hazardous Waste Report respondent name
RESPNDTX	A8 1993 Hazardous Waste Report respondent name (preloaded)
RESPPHON	A8 1993 Hazardous Waste Report respondent phone number
RESPPHOX	A8 1993 Hazardous Waste Report respondent phone number (preloaded)
RESPTITL	A8 1993 Hazardous Waste Report respondent title
RESPTITX	A8 1993 Hazardous Waste Report respondent title (preloaded)
SEQ	Used for ordering comments records when printing comments (all values are 0 in the FACILTYx file)
TRID	A5 Facility TRI number
WSDATA	A flag (Yes or No) indicating whether the facility responded to the survey by providing data for at least one waste stream

WSNUM	Waste stream number used for ordering comments by waste stream when printing comments (all values are 0 in the FACILTYx file)

File name: WSTREAMx

Variable name	Description
AGGWGT	Sampling weight for the selection of one of several similar (aggregate) waste streams
ASHBASIS	B16 _4 Percent ash by weight: Basis for estimate (analysis of more than 5 samples or estimate from less data, plus year from which data was used to estimate value for 1993, if not 1993)
BIO5BASI	B16_12B BOD5 Basis for estimate (analysis of more than 5 samples or estimate from less data, plus year from which data was used to estimate value for 1993, if not 1993)
BIO5DAYA	B16_12A 5-day Biological oxygen demand: qualifier (NA, NT, <, >, or blank)
BIO5DAYB	B16_12B 5-day Biological oxygen demand: average or low end of range
BIO5DAYC	B16_12C 5-day Biological oxygen demand: range indicator (- or blank)
BIO5DAYD	B16_12D 5-day Biological oxygen demand: upper end of range (if any)
BIO5UNIT	B16_12U 5-day Biological oxygen demand: unit of measure (#/DAY, %, MG/L, PPM, or blank)
BRS_PAGE	Page on which the waste stream was reported, as recorded in the BRS
BRS_SUBP	Subpage on which the waste stream was reported, as recorded in the BRS
BRSDENSX	Density of waste stream as reported in the 1993 BRS
BRSDUOMX	Unit of measure for the waste stream density as reported in the 1993 BRS
BRSWGT	1=BRS reportable waste, 0=Waste which did not need to be reported to BRS
BTUA	B16_10A BTU content: qualifier (NA, ND, <, >, NT, or blank)
BTUB	B16_10B BTU content: average or low end of range
BTUBASIS	B16_10B BTU content: Basis for estimate (analysis of more than 5 samples or estimate from less data, plus year from which data was used to estimate value for 1993, if not 1993)
BTUC	B16_10C BTU content: range indicator (- or blank)
BTUD	B16_10D BTU content: upper end of range (if any)
BTUUNIT	B16_10U BTU content: unit of measure (/GAL, /LB, or blank)
C2	C2 Point of measurement code for constituent concentrations
CBRSSPGR	BRS density (BRSDENSX) converted to specific gravity
CHEMOXA	B16_13A Chemical oxygen demand (COD): qualifier (NA, NT, <, >, or blank)
CHEMOXB	B16_13B Chemical oxygen demand (COD): average or low end of range
Variable name	Description
CHEMOXC	B16_13C Chemical oxygen demand (COD): range indicator (- or blank)
CHEMOXD	B16_13D Chemical oxygen demand (COD): upper end of range (if any)
CHOXBASI	B16_13B Chemical oxygen demand (COD): Basis for estimate (analysis of more than 5 samples or estimate from less data, plus year from which data was used to estimate value for 1993, if not 1993)
CHOXUNIT	B16_13U Chemical oxygen demand (COD): unit of measure (%, MG/L, PPM, or blank)
CONSTNUM	Constituent number used for ordering comments by constituent when printing comments(all values are 0 in the WSTREAMx file)
CSPECGRA	Specific Gravity/Density converted to Specific Gravity
CSYQTY1X	Preloaded quantity 1 converted to Tons
CSYQTY2X	Preloaded quantity 2 converted to Tons

CSYQTY3X	Preloaded quantity 3 converted to Tons
CSYQTY4X	Preloaded quantity 4 converted to Tons
CSYSQTY	Sum of actual quantities managed (1-5) converted to Tons
CSYSQTY1	SYSTDR1 actual quantity 1 converted to Tons
CSYSQTY2	SYSTDR2 actual quantity 2 converted to Tons
CSYSQTY3	SYSTDR3 actual quantity 3 converted to Tons
CSYSQTY4	SYSTDR4 actual quantity 4 converted to Tons
CSYSQTYX	Sum of preloaded quantities (1-5) converted to Tons
EPAID	EPA Facility ID
FLASHA	B16_9A Flash point: qualifier (NA, NT, <, >, or blank)
FLASHB	B16_9B Flash point: average or low end of range
FLASHC	B16_9C Flash point: range indicator (- or blank)
FLASHD	B16_9D Flash point: upper end of range (if any)
FLSHBASI	B16_9B Flash point: Basis for estimate (analysis of more than 5 samples or estimate from less data, plus year from which data was used to estimate value for 1993, if not 1993)
FLSHUNIT	B16_9U Flash point: unit of measure (F, C, or blank)
FORMCODE	B1 Waste form code
FORMCODX	B1 Waste form code (preloaded)
FSYSQTYX	Flag indicating that SYSQTYX and SYSQTY (where applicable) have been modified (01 = change)
FSYSTDR1	A flag indicating if the respondent changed SYSTDR1 (-1 = change)
FSYTDR1X	A flag indicating if the respondent changed SYSTDR1X (-1 = change)
Variable name	Description
FWSTCODC	Flag indicating that the original waste codes have been modified (-1 = change). Changes include removing comments and waste codes.
FWSTCODF	Flag indicating WSTCODES has been modified (-1 = change) (filled waste streams with pretest data).
FWSTCODX	A flag indicating if the respondent provided a change to the waste codes WSTCODEX (-1 = change)
GM_WR	BRS form on which the waste stream was reported (Form GM or Form WR)
H20BASIS	B16_5B Percent water by weight: Basis for estimate (analysis of more than 5 samples or estimate from less data, plus year from which data was used to estimate value for 1993, if not 1993)
HALOBASI	B16_11B Total organic halogen (percent by weight): Basis for estimate (analysis of more than 5 samples or estimate from less data, plus year from which data was used to estimate value for 1993, if not 1993)
HALOGENA	B16_11A Total organic halogen (percent by weight): qualifier (NA, ND, <, >, TR, NT, or blank)
HALOGENB	B16_11B Total organic halogen (percent by weight): average or low end of range
HALOGENC	B16_11C Total organic halogen (percent by weight): range indicator (- or blank)
HALOGEND	B16_11D Total organic halogen (percent by weight): upper end of range (if any)
HASCCON	A flag (Y or N) indicating if the respondent provided positive constituent concentrations for at least one constituent in the one waste stream. N if the waste stream was not managed on-site.
HASCDAT	A flag (Y or N) indicating if the respondent provided any constituent concentration information for at least one constituent in the one waste stream.
INSURWGT	1=Information about the waste stream was requested as part of the survey, 0=

	Information about the waste stream was volunteered and not requested
NHWCSSIC	Combination of BRS and RCRIS/FINDS SIC Codes (if waste from GM then use SICCODE, if waste from WR then use RCFNSIC
OFFEPAID	B11 The EPA Facility ID of the off-site facility from which the waste was received
OFFEPAIX	B11 The EPA Facility ID of the off-site facility from which the waste was received (preloaded)
OILBASIS	B16_7B PCT OIL BY WT Basis for estimate (analysis of more than 5 samples or estimate from less data, plus year from which data was used to estimate value for 1993, if not 1993)
ORIGCODE	B7 Origin code
ORIGCODX	B7 Origin code (preloaded)
Variable name	Description
ORIGWCOD	Reported waste codes (WSTCODES) before modifications
PCTASHA	B16_4A Percent ash by weight: qualifier (NA, ND, <, >, NT, or blank)
PCTASHB	B16_4B Percent ash by weight: average or low end of range
PCTASHC	B16_4C Percent ash by weight: range indicator (- or blank)
PCTASHD	B16_4D Percent ash by weight: upper end of range (if any)
PCTBASI	B16_2B Percent solids by weight: Basis for estimate (analysis of more than 5 samples or estimate from less data, plus year from which data was used to estimate value for 1993, if not 1993)
PCTDRVFR	B14 Percentage of waste stream consisting of "derived from" hazardous waste
PCTNOHAZ	B15 Percentage of waste stream consisting of non-hazardous waste
PCTOILA	B16_7A Percent oil by weight: qualifier (NA, ND, <, >, TR, NT, or blank)
PCTOILB	B16_7B Percent oil by weight: average or low end of range
PCTOILC	B16_7C Percent oil by weight: range indicator (- or blank)
PCTOILD	B16_7D Percent oil by weight: upper end of range (if any)
PCTSOLIA	B16_2A Percent solids by weight: qualifier (NA, ND, <, >, NT, or blank)
PCTSOLIB	B16_2B Percent solids by weight: average or low end of range
PCTSOLIC	B16_2C Percent solids by weight: range indicator (- or blank)
PCTSOLID	B16_2D Percent solids by weight: upper end of range (if any)
PCTTOCA	B16_6A Percent total organic carbon (TOC) by weight: qualifier (NA, NT, <, >, or blank)
PCTTOCB	B16_6B Percent total organic carbon (TOC) by weight: average or low end of range
PCTTOCC	B16_6C Percent total organic carbon (TOC) by weight: range indicator (- or blank)
PCTTOCD	B16_6D Percent total organic carbon (TOC) by weight: upper end of range (if any)
PCTTSSA	B16_3A Percent total suspended solids (TSS) by weight: qualifier (NA, ND, <, NT, or blank)
PCTTSSB	B16_3B Percent total suspended solids (TSS) by weight: average or low end of range
PCTTSSC	B16_3C Percent total suspended solids (TSS) by weight: range indicator (- or blank)
PCTTSSD	B16_3D Percent total suspended solids (TSS) by weight: upper end of range (if any)
PCTWATEA	B16_5A Percent water by weight: qualifier (NA, ND, <, >, TR, NT, or blank)
Variable name	Description
PCTWATEB	B16_5B Percent water by weight: average or low end of range
PCTWATEC	B16_5C Percent water by weight: range indicator (- or blank)

PCTWATED	B16_5D Percent water by weight: upper end of range (if any)
PHA	B16_8A PH qualifier (NA, ND, <, >, NT, or blank)
PHB	B16_8B PH average or low end of range
PHBASIS	B16_8B PH Basis for estimate (analysis of more than 5 samples or estimate from less data, plus year from which data was used to estimate value for 1993, if not 1993)
PHC	B16_8C PH range indicator (- or blank)
PHD	B16_8D PH upper end of range (if any)
PHUNIT	B16_8U PH unit of measure
PHYSCHAR	B17 Code for point of measure for the waste stream characteristics
POMCODE	B8 Code for the point of measure for the volume of waste streams generated on-site (see QUAN1993)
POMCODEX	B8 Code for the point of measure for the volume of waste streams generated on-site (see QUAN199X, preloaded)
QGENUNIT	B10 Code for unit of measure of the quantity generated (QUAN1993)
QGENUNIX	B10 Code for unit of measure of the quantity generated (QUAN199X, preloaded)
QORDER	Used for ordering comments by question within section when printing comments (all values are 0 in the WSTREAMx file)
QRECUNIT	B13 Unit of measure for the quantity of waste received from off-site (QUANRECV)
QRECUNIX	B13 Unit of measure for the quantity of waste received from off-site (QUANRECX, preloaded)
QUAN1993	B9 Quantity of waste generated on-site
QUAN199X	B9 Quantity of waste generated on-site (preloaded)
QUANRECV	B12 Quantity of waste received from off-site
QUANRECX	B12 Quantity of waste received from off-site (preloaded)
RCFNSIC	RCRIS/FINDS SIC Codes (Generated and Managed)
RCRAEXMP	B3 'Y" indicates waste managed exclusively in units exempt from RCRA permitting
RCRAEXMX	B3 'Y" indicates waste managed exclusively in units exempt from RCRA permitting (preloaded)
SAMPWGT	Sampling weight for the selection of 20 waste streams if there were more than 20 large waste streams at the facility
Variable name	Description
SEQ	Used for ordering comments records when printing comments (all values are 0 in the WSTREAMx file)
SIC	SIC code for the waste generator (SIC code for the respondent facility, if the generator of the waste), otherwise, SIC code for the facility from which the waste was received (obtained from RCRIS/FINDS or BRS for other years).
SICCODE	B5 SIC code for the process which generated the waste, either preloaded from BRS or as reported by the respondent
SICCODEX	B5 SIC code for the process which generated the waste (preloaded)
SPECGRA	Specific Gravity/Density converted from SPECGRAA through SPECGRAD (A through D) using the following algorithm:

If A = "NA", "ND", "NT", "TR" then SPECGRA = 0
If A = "<" then SPECGRA = B/2
If A = ">" then SPECGRA = B
If C = "-" then SPECGRA = (B+D)/2
Else SPECGRA = B

SPECGRAA	B16_1A Specific gravity or density: qualifier (NA, NT, >, or blank)

SPECGRAB	B16_1B Specific gravity or density: average or low end of range
SPECGRAC	B16_1C Specific gravity or density: range indicator (- or blank)
SPECGRAD	B16_1D Specific gravity or density: upper end of range (if any)
SPGRBASI	B16_1B Specific gravity or density: Basis for estimate (analysis of more than 5 samples or estimate from less data, plus year from which data was used to estimate value for 1993, if not 1993)
SPGRNEW	Specific Gravity converted from CSPECGRA and CBRSSPGR using the following algorithm:
If CSPECGRA is between 0.5 and 5 then used SPECGRA, else if CBRSSPGR is between 0.5 and 5 then use CBRSSPGR, else use 1.0	
SPGRUNIT	B16_1U Specific gravity or density: unit of measure (#/FT, #/GAL, #/Y3, G/ML, g/ml, T/Y3, or blank)
SRCCODE	B6 Source code
SRCCODEX	B6 Source code (preloaded)
STATE_CD	State waste codes (if any)
STATE_CX	State waste codes (if any, preloaded)
SYSQTY	Sum of actual quantities managed (1-5)
SYSQTY1	B4 Quantity of waste managed in the first system type
SYSQTY1X	B4 Quantity of waste managed in the first system type (preloaded)
SYSQTY2	B4 Quantity of waste managed in the second system type
SYSQTY2X	B4 Quantity of waste managed in the second system type (preloaded)
SYSQTY3	B4 Quantity of waste managed in the third system type
Variable name	Description
SYSQTY3X	B4 Quantity of waste managed in the third system type (preloaded)
SYSQTY4	B4 Quantity of waste managed in the fourth system type
SYSQTY4X	B4 Quantity of waste managed in the fourth system type (preloaded)
SYSQTY5	B4 Quantity of waste managed in the fifth system type
SYSQTY5X	B4 Quantity of waste managed in the fifth system type (preloaded)
SYSQTYX	Sum of preloaded quantities (1-5)
SYSTDR1	B4 System type code for the first system type
SYSTDR1X	B4 System type code for the first system type (preloaded)
SYSTDR2	B4 System type code for the second system type
SYSTDR2X	B4 System type code for the second system type (preloaded)
SYSTDR3	B4 System type code for the third system type
SYSTDR3X	B4 System type code for the third system type (preloaded)
SYSTDR4	B4 System type code for the fourth system type
SYSTDR4X	B4 System type code for the fourth system type (preloaded)
SYSTDR5	B4 System type code for the fifth system type
SYSTDR5X	B4 System type code for the fifth system type (preloaded)
TOCBASIS	B16_6B Percent total organic carbon (TOC) by weight: Basis for estimate (analysis of more than 5 samples or estimate from less data, plus year from which data was used to estimate value for 1993, if not 1993)
TSSBASIS	B16_3B Percent total suspended solids (TSS) by weight: Basis for estimate (analysis of more than 5 samples or estimate from less data, plus year from which data was used to estimate value for 1993, if not 1993)
UNIT	Unit of actual quantities managed (1-5)
UNIT1X	B4 Unit of measure for quantity managed in the first system type (preloaded)
UNIT2X	B4 Unit of measure for quantity managed in the second system type (preloaded)
UNIT3X	B4 Unit of measure for quantity managed in the third system type (preloaded)
UNIT4X	B4 Unit of measure for quantity managed in the fourth system type

	(preloaded)
UNIT5X	B4 Unit of measure for quantity managed in the fifth system type (preloaded)
UNITS1	B4 Unit of measure for quantity managed in the first system type
UNITS2	B4 Unit of measure for quantity managed in the second system type
UNITS3	B4 Unit of measure for quantity managed in the third system type
UNITS4	B4 Unit of measure for quantity managed in the fourth system type
Variable name	Description
UNITS5	B4 Unit of measure for quantity managed in the fifth system type
UNITX	Unit of preloaded quantities (1-5)
WSDAT	A flag (Yes or No) indicating whether the facility responded to the survey by providing data this waste stream
WSNUM	Waste stream number, sequential number identifying each waste stream within a facility
WST_ORG	Waste origin code
WST_ORGX	Waste origin code (preloaded)
WSTCODES	EPA waste codes for the waste stream
WSTCODEX	EPA waste codes for the waste stream (preloaded)
WSTDESC	Waste description
WSTDESCX	Waste description (preloaded)
WSTYPE	Waste stream type: indicate special circumstances surrounding the waste stream data in the data files
WW	B2 "Y" indicates that the waste stream is a waste water
WWX	B2 "Y" indicates that the waste stream is a waste water as indicated by information in the BRS

File name: CONSTITx

Variable name	Description
CAS	C1 Chemical Abstract Service (CAS) number for the constituent, revised based on a review of the information provided by the respondents (CASR). Contains valid CAS numbers in the form #####-##-#, where # is a digit. Blank indicates that no valid CAS number was found. Otherwise:

No CASRN = no CAS number for the constituent was believed to exist;
SEQ: 25 = Alkyl benzenes;
N230 = Glycol ethers.

CASR	C1 Chemical Abstract Service (CAS) number for the constituent as reported by respondent
CASX	C1 Chemical Abstract Service (CAS) number for the preloaded constituents
CHEMMASS	Chemical mass in the waste stream in tons using the following algorithm:

CHEMMASS = Quantity(Tons) [CSYSQTY]* Concentration(PPM)/1000000
where Concentration(PPM) is the whole waste concentration (CWAVGVAL) if it is greater than zero, otherwise, the leachate concentration (CLAVGVAL) times 20

CLAVGVAL	The leachate concentration (LAVGVAL) converted to ppm by weight.
CONSTIT	C1 Constituent name, revised based on a review of the information provided by the respondents. CONSTIT = CONSTITR if no valid CAS number and corresponding constituent name was found.
CONSTITR	C1 Constituent name as reported by respondent
CONSTITX	C1 Preloaded constituent name for the preloaded constituents
CONSTNUM	Constituent number, sequential number identifying each constituent within each waste stream

Variable name	Description
CSOURCE	Indicates the source of the revised CAS-constituent name: Preferred = CASR and CONSTITR matched correct values and CONTITR was the preferred name; Okname = CASR and CONSTITR matched correct values and CONTITR was an accepted name for the constituent; ALTname = After correcting for spelling errors, CASR and CONSTITR matched correct values; Camatsd or Camatsc = CASR matched a valid CAS number and CONSTITR was similar to the valid constituent name for the CAS number. CONSTIT was set to the constituent name for the valid CAS. Two names were judged similar if CONSTITR was a substring within the valid constituent name or the valid constituent name was a substring within CONSTITR. Camat = CASR matched a valid CAS number but CONSTITR was not similar (as defined above) to the valid constituent name for the CAS number. After review these matches were all accepted. When the CAS number and constituent name were both valid but not consistent with each other, a match on CAS number was accepted as correct; Comatbl = CONSTITR matched a valid constituent name and CASR was blank. CAS was set to the CAS number for the valid constituent name. Comatbl2 = CONSTITR matched a valid constituent name, after correcting for apparent spelling errors, and CASR was blank. CAS was set to the CAS number for the valid constituent name. Comatch = CONSTITR matched a valid constituent name but CASR was not valid. After review, the matche as accepted and CAS was set to the CAS number for the valid constituent name. Comatch2 = CONSTITR matched a valid constituent name after correcting for spellin errors but CASR was not valid. After review, the match was accepted and CAS was set to the CAS number for the valid constituent name. nomatch = neither CASR nor CONSTITR matched a correct value, CAS is blank, CONSTIT = CONSTITR;
CWAVGVAL	The whole waste concentration (WAVGVAL) converted to ppm by weight.
EPAID	EPA facility ID
FFIXDUP	Flag indicating that CWAVGVAL, CLAVGVAL, and CHEMMASS were adjusted to remove concentrations and chemical masses for constituents with duplicate records in a waste stream. The rules for removing duplicates chose the record with the highest PPM concentration (CWAVGVAL or CLAVGVAL). Other values were set to missing. FFIXDUP = 1 indicates a record for which CWAVGVAL, CLAVGVAL, and CHEMMASS were set to missing. FFIXDUP = 2 indicates a record with the highest whole waste concentration (CWAVGVAL) for which the leachate concentration (CLAVGVAL) came from a different record.

Variable name	Description
LAVGVAL	Leachate concentration determined from LAVGVALA through LAVGVALD (A through D) using the following algorithm:

If A = "NA", "ND", "NT", "TR" then WAVGVAL = 0
If A = "<" then WAVGVAL = B/2

If A = ">" then WWAVGVAL = B
If C = "-" then WAVGVAL = (B+D)/2
Else WAVGVAL = B.

LAVGVALA	Leachate concentration of the constituent: qualifier (NA, ND, <, >, DK, NT, PR, or blank)
LAVGVALB	Leachate concentration of the constituent: average or low end of range
LAVGVALC	Leachate concentration of the constituent: range indicator (- or blank)
LAVGVALD	Leachate concentration of the constituent: upper end of range (if any)
LCBASIS	Leachate concentration of the constituent: Basis for estimate (analysis of more than 5 samples or estimate from less data, plus year from which data was used to estimate value for 1993, if not 1993)
MODLPPM	Contains the ratios by which the reported leachate concentrations were multiplied to obtain a total of 100% in those waste streams which were adjusted (those with totals exceeding 110%).
MODWPPM	Contains the ratios by which the reported whole waste concentrations were multiplied to obtain a total of 100% in those waste streams which were adjusted (those with totals exceeding 110%).
LCTYPE	A character variable showing the form of the leachate constituent concentration data: #, #-#, 0, <#, <#-#, >#, DK,NA,ND, NT, PR, TR, or blank, where # indicates a numeric value.
LCUNIT	Leachate concentration of the constituent: unit of measure (%, MG/L, PPB, PPM, UG/L, V%, or blank)
QORDER	Used for ordering comments by question within section when printing comments (all values are 0 in the CONSTITx file)
SEQ	Used for ordering comments records when printing comments (all values are 0 in the CONSTITx file)
WAVGVAL	Whole waste concentration determined from WAVGVALA through WAVGVALD (A through D) using the following algorithm:

If A = "NA", "ND", "NT", "TR" then WAVGVAL = 0
If A = "<" then WAVGVAL = B/2
If A = ">" then WWAVGVAL = B
If C = "-" then WAVGVAL = (B+D)/2
Else LAVGVAL = B.

Variable name	Description
WAVGVALA	Concentration of the constituent in the whole waste: qualifier (NA, ND, <, >, DK, NT, PR, TR, or blank)
WAVGVALB	Concentration of the constituent in the whole waste: average or low end of range
WAVGVALC	Concentration of the constituent in the whole waste: range indicator (- or blank)
WAVGVALD	Concentration of the constituent in the whole waste: upper end of range (if any)
WSNUM	Waste stream number, sequential number identifying each waste stream within a facility
WWBASIS	Concentration of the constituent in the whole waste: Basis for estimate (analysis of more than 5 samples or estimate from less data, plus year from which data was used to estimate value for 1993, if not 1993)
WWTYPE	A character variable showing the form of the whole waste constituent concentration data: #, #-#, 0, <#, <#-#, >#, DK,NA,ND, NT, PR, TR, or blank, where # indicates a numeric value.
WWUNIT	Whole waste concentration of the constituent: unit of measure (%, MG/L,

PPB, PPM, UG/L, V%, or blank)

File name: **COMMENTx**

Variable name	Type	Length	Description
COMMENT			Comment text
CONSTNUM			Constituent number indicating the constituent to which the comment applies (zero if the comment is for facility or waste stream information)
EPAID			EPA facility ID
QORDER			Used for ordering comments by question within section when printing comments
QUES			Number of the question to which the comment applies
SEQ			Indicates the order of comment records which have the same EPAID, WSNUM, CONSTNUM, and QUES
WSNUM			Waste stream number indicating the waste stream to which the comment applies (zero if the comment is for facility information)

Appendix 5

Westat received computer disks with constituent data from four Rollins Corporation facilities. The data was provided to EPA in association with facility visits by EPA personnel. Although the data have some of the information contained in the NHWCS data files, the methods of data collection, documentation of methods, and format of the Rollins data differs considerably from the NHWCD data. Therefore, the Rollins data are provided in a separate file rather than as part of the NHWCS data files. The preparation of the Rollins data files are described below.

Because the original data files were poorly documented, preparing the data required making some assumptions about the format and meaning of the data. Those assumptions are described below. We have reorganized the data to make them more usable. All of the original data is included in these files. None of the data have been changed.

The data came to Westat in four compressed files and three files which apparently contained the file layouts. The file layouts indicated that the files had fields enclose by quotes and separated by commas. The file layout also indicated that the files had a variable describing the process that generated the waste. As a result of looking at the files, it appeared that the process description was not present, that fields were separated by commas, that non-numerical fields were enclosed by quotes. In a few records there were fewer fields than expected. After looking at the file contents, we assumed that the missing fields were fields at the end of the records and were empty.

There were three types of files indicated by the letters "cas", "epa", and "chem" in the file names. Each file apparently contained the waste stream number, quantity of waste (year to date and prior yr), physical form of the waste (powder, gas, sludge, liquid, and solid), ash content, heat of combustion, and "scrub" value. In addition, the "cas" files apparently contained a cas registry number, a value (perhaps a concentration), and a quantity (pounds). The "chem" files apparently contained a chemical symbol for an element

(usually a metal) and a concentration. The "epa" files apparently contained EPA waste codes for the waste stream. The best information on the contents of the fields is presented below.

Numerical values are often described by three variables, an indicator (such as = or <) a low value and a high value. Both the low and high values may be zero. If the low value is greater than zero, the high value is often zero. If both the low and high values are present and non-zero, the indicator is often "R" suggesting that the number represent a range. However, there are roughly 100 cases where the low value is greater than the high value and both are greater than zero.

It appeared that the location of the Rollin's facility and the waste stream number could be used to link the information in the different files. We checked that the information on the quantity of waste, physical form, ash content, heat of combustion, and "scrub" value were the same in all records with the same waste stream number in all files. These values were the same except for the quantities of waste. On the assumption that waste stream was a unique waste stream identifier within a facility and that the quantities of waste were occasionally different across files for unknown reasons, the EPA facility ID was added to the files and the EPA ID and waste stream number can be used to link all files.

The files are organized as follows:

ROL_WS contains the waste stream information (similar to the waste stream files for the survey data).

ROL_EPA has the EPA waste codes for the waste streams, with one waste code per record and usually multiple records per waste stream. There were as many as 502 waste codes per waste stream so we did not try to put them into one variable. None of the waste streams from the Louisiana facility have waste codes. There are many other waste streams that have no EPA waste codes.

ROL_CAS contains what is apparently information about different (non-element) constituents. There may be multiple records per waste stream.

ROL_CHEM contains information about elements (usually metals) in the waste stream. There may be multiple records per waste stream.

The files are available in compressed form as both SAS files (version 6.11 with the extension .SD2) and ASCII files (with the extension .DAT and an associated dictionary file). In these files, the numerical values were stored as character values.

The following descriptions of the variables in the files are based on the description provided in the file layout and on an analysis of the data in the files.

File name: ROL_WS, 16,074 records

Variable name	Description in file layout	Comments
EPAID	The EPA facility ID, either LAD010395127, NJD053288239, or TXD055141378. Determined from the file name and other documentation.	
STREAMN	stream nbr	Apparently a waste stream number provided by Rollins
ASHPHIGH	ash pct low	A number from 0 to 10,000
ASHPIND	ash pct indicator	Missing or one of: < = > A R
ASHPLOW	ash pct high	A number from 0 to 10,000
GAS	gas	Missing or one of: Y N
HOCHIGH	heat of combustion high	A number from 0 to 62,000 or missing
HOCIND	heat of combustion indicator	Missing or one of: < = > A R
HOCLOW	heat of combustion low	A number from 0 to 80,000 or missing
LIQUID	liquid	Missing or one of: Y N
NCAS	The number of records in the ROL_CAS file for this waste stream (missing or 1 up to 40)	
NCHEM	The number of records in the ROL_CHEM file for this waste stream (missing or 1 up to 89)	
NEPA	The number of records in the ROL_EPA file for this waste stream (missing or 1 up to 502)	
POWDER	powder	Missing or one of: Y N
QPREVCAS	pounds prior yr	A number from 0 to 6,909,070 or missing
QPREVCHM	pounds prior yr	A number from 0 to 6,909,070 or missing
QPREVEPA	pounds prior yr	A number from 0 to 6,909,070 or missing
QYTDCAS	pounds ytd	A number from 0 to 1,226,100 or missing
QYTDCHM	pounds ytd	A number from 0 to 1,226,950 or missing
QYTDEPA	pounds ytd	A number from 0 to 1,107,450 or missing
SCRBHIGH	scrub high	A number from 0 to 2,000
SCRBIND	scrub indicator	Missing or one of: < = > A R
SCRBLOW	scrub low	A number from 0 to 5,000
SLUDGE	sludge	Missing or one of: Y N
SOLID	solid	Missing or one of: Y N

File name: ROL_EPA, 39,738 records

Variable name	Description in file layout	Comments
EPAID	The EPA facility ID, either LAD010395127, NJD053288239, or TXD055141378.	
STREAMN	stream nbr	Apparently a waste stream number provided by Rollins
EPACODE	epa code	missing or a code of the form ANNNXX where A stands for a letter in: C D F K P X. NNN stands for a three digit number and XX stands for a two character suffix, if present.

File name: ROL_CHEM, 464,295 records

Variable name	Description in file layout	Comments
EPAID	The EPA facility ID, either LAD010395127, NJD053288239, or TXD055141378.	
STREAMN	stream nbr	Apparently a waste stream number provided by Rollins
CHEMHIGH	chemical high	A number from 0 to 1,000,000,000
CHEMIND	chemical indicator	Missing or one of: / 0 < = > A P R
CHEMLOW	chemical low	A number from 0 to 2,953,000,000
CHEMSYMB	chemical symbol	A one or two digit chemical symbols (such as Ag, Hg,

Se). We did not check if all codes were legitimate elements.
CHEMUOM chemical unit of measure missing or one of: 000000 00{000 0{0000 CT MG MH/KG MG/L PCT PPB PPH PPM mg/l ppm

File name: ROL_CAS, 66,377 records

Variable name	Description in file layout	Comments
EPAID	The EPA facility ID, either LAD010395127, NJD053288239, or TXD055141378.	
STREAMN	stream nbr	Apparently a waste stream number provided by Rollins
CASLBS	cas registry pounds	A number from 0 to 999,000
CASVALUE	cas registry value	A number from 0 to 6,287,254
CASNUM	cas registry number	An 11 digit number with leading zeros preceded by either the letter "C" or "R". One value starts with a "V". for example: C00000075854.