

US EPA ARCHIVE DOCUMENT

Overview of the Electronic Data Files for USEPA's 1996 National Hazardous Waste Constituent Survey (NHWCS) Database

I. BRIEF INTRODUCTION TO THE NHWCS

This overview is to provide an introduction to initial users of the USEPA Office of Solid Waste's electronic copy of the 1996 "National Hazardous Waste Constituent Survey" (NHWCS) database.

This 1996 voluntary participation survey attempted to collect data about the physical characteristics and chemical constituents (identity and concentrations) of industrial hazardous wastestreams, from a sample consisting of 221 of the nation's largest hazardous waste treatment, disposal and recycling facilities (TDRFs), including nine facilities which participated in a prior survey pretest. Non-CBI survey responses were received from 156 of the sample facilities, plus one facility which claimed CBI status and thus was not included in the resultant database for public release.

At a comprehensive level, the NHWCS database may be described as consisting of 11 separate components (i.e. electronic files), which may be grouped into **three categories**:

- **Introductory Documentation Files**: Three files which contain descriptive documentation and summary information for initial users (e.g. one file provides a 24-slide overview of the NHWC Survey, and other files define datafields in the database).
- **Data Summary Files**: Three files which summarize the data contained in the database, consisting of an OSW contractor summary report, and two spreadsheet files which present summary data.
- **Raw Data Files (Database)**: Five components (electronic files) which actually contain the computerized raw data from the survey questionnaires.

These **11 files** total **14.7 MB** and are listed and described separately below. To facilitate Internet transfer and downloading, these 11 files are "zipped" into three separate files. All files may be transferred between different over-the-counter software programs, but are provided by OSW for user access convenience, in the following electronic formats: (a) Paradox (*.db); (b) Lotus-123 (*.wk4); (c) WordPerfect (*.wpd or *.doc), (d) ascii text (*.txt), and (e) PowerPoint (*.ppt).

II. INTRODUCTORY DOCUMENTATION FILES (n=3 files, 1.317 MB total size)

The three different files listed below present different levels of details as introductions to the survey and the resultant database. These three files are zipped together as a single file “**zipintro.ZIP**”:

- | The intended starting point for initial users is to access and read the present three-page overview (“**overview.wpd**” , 19 KB), which briefly defines each component of the NHWCS database.
- | A PowerPoint file (“**briefing.ppt**”, 930 KB) provides a convenient presentation-style overview, consisting of a 24-slide, OSW briefing package from 26 May 1998.
- | A text file (“**datadoc5.doc**”, 368 KB) written by an OSW contractor (Westat, Inc., dated 30 March 1998), presents 52 pages of detailed, descriptive information about the assorted NHWCS datafields in each of the five database component files (defined below), and describes the scope and methodology of the survey design. This file should also be viewed by initial users before attempting to access the raw data files.

III. DATA SUMMARY FILES (n=3 files; 6.53 MB total size)

Three files provide the user with descriptive and summary data and information about the design and findings of the NHWCS. These three files contain summary data extracted by OSW contractors from the NHWCS database. The first file listed below – a seven chapter, 63-page report which summarizes highlights from the database – may also be instructive to initial users, and used as a reference guide to the types of aggregative analyses which may be performed using the NHWCS database. Although listed separately, these three files are zipped together as a single file “**zip_smry.ZIP**”:

- | A WordPerfect-8 electronic file (“**summary.wpd**”, 5.28 MB), contains OSW-EMRAD's contractor's final (October 1998) summary report, which describes the main features, elements, scope and major findings of the NHWCS database (complete with data tables and graphs which aggregate numerous datafields).
- | An OSW-contractor's Lotus-123 file (“**chemsum.wk4**”, 1.19 MB) contains multiple spreadsheets with summarize the database information, according to key datafields for about +/-740 waste constituents (with and without RCRA-exempt TSDR units).
- | A third summary type file (“**cooccur.wk4**”, 60 KB) provides a frequency count of the co-occurrence of waste constituent in wastestreams reported in the database. The file contains two spreadsheets with this information.

IV. RAW SURVEY DATA FILES (n=5 files; 6.88 MB total size)

Also attached are a suite of five data files all with the extension *.db (Corel Paradox-8 extension), which are the five separate pieces or components which actually constitute the raw data of OSW's 1996 NHWCS database. That is, the NHWCS database actually consists of five separate databases, all of which may be electronically interlinked (joined) together, by using the three datafields: (a) facility EPA ID number (EPAID), (b) wastestream ID number (WSNUM), or (c) constituent number (CONSTNUM).

The facility, wastestream, and constituent datafiles contain two types of information: the preloaded information (derived from the 1993 BRS) provided on the questionnaire, and the data provided by the respondents. If the respondents did not change the preloaded information, the datafields with the respondent information show the preloaded data.

Each of these five component databases can be conceptualized/visualized as representing data matrices consisting of horizontal matrix rows (i.e. data records or entries), and vertical matrix columns (i.e. datafields or data elements). NHWCS contains a non-duplicative total of 71 datafields in all five data files (although some datafields such as facility EPA ID are contained in more than one component data file).

The five electronic components by *.db filename of OSW's NHWCS database, and associated number of matrix rows and datafield columns are listed separately below; they are zipped together as a single file "**zip_data.ZIP**":

- | "**facility.db**" (66 KB) = A total of 221 large quantity industrial hazardous waste treatment, disposal and recycling facilities (rows or records), containing 13 datafields (i.e. data matrix columns).
- | "**wastestr.db**" (1.18 MB) = 1,760 wastestreams (rows or records), containing 25 datafields (this file contains SIC codes for the process which generated the waste, annual waste volume, RCRA wastecodes, etc.).
- | "**constit.db**" (5.23 MB) = 36,609 waste constituent records, containing ten datafields (this file contains waste constituent concentration and mass data).
- | "**physchar.db**" (238 KB) = physical characteristics data for each of the 1,760 wastestreams, containing 18 datafields.
- | "**comment.db**" (166 KB) = 892 open-ended written comments for wastestreams, containing five datafields.

Although these five database files are formatted with the extension *.db (Corel Paradox-8) files, you should be able to transfer and access them using other database software programs, such as D-Base, SAS, MicroSoft-Access, SPSS, Stata, or spreadsheet software such as MicroSoft-Excel or Lotus-123 (although the "constit.db" file which contains a data matrix consisting of 36,609 rows, exceeds Lotus' 1994 spreadsheet limit of 8,192 rows per spreadsheet).

If you have any follow-up questions or difficulty with these accessing and interpreting the data and information contained in these files, contact USEPA-OSW at 703-308-8615, or via the "Comments" weblink at the Internet website <http://www.epa.gov/epaoswer> .

