



## *Project Summary*

# The Revised Organic Chemical Producers Data Base System

G. E. Wilkins, C. H. Tucker, and E. D. Gibson

This report describes the revised Organic Chemical Producers Data Base (OCPDB), an automated chemical industry information system developed in 1976 for the U.S. Environmental Protection Agency (EPA).

Improvements by Radian Corporation, Austin, Texas, have been made in two ways: (1) expansion of the data base to include more chemicals and more information about each chemical, and (2) implementation of the system through a data base management system.

The revised data base includes almost 600 chemicals and their more than 1300 producers. Chemicals are described by Chemical Abstracts Services (CAS) registry number, Wiswesser Line Notation (WLN), industrial process descriptions, chemical uses, synonyms, toxicity data, economic data, and producers. Priority pollutants identified as a result of Natural Resources Defense Council (NRDC) vs. EPA are marked and process descriptions are cross-referenced with another EPA reference source, The Industrial Process Profiles for Environmental Use (IPPEU) Chapter 6. Locations of producers are described by city, state, EPA region, and river basin. The chemicals that are produced at each location are listed, along with name-plate capacities, when available.

Retrieval is possible through use of any of a number of "key" data elements: chemical name, synonyms,

OCPDB number, CAS numbers, WLN, priority pollutant markers, process ID number, IPPEU numbers, producer company name, parent company name, city, state, river basin, and EPA region.

### Introduction

This report describes the revised Organic Chemical Producers Data Base (OCPDB) system. The original OCPDB was developed in 1976 for EPA's Industrial Environmental Research Laboratory (IERL) in Cincinnati under EPA Contract 68-02-1319, Task 51.

The computerized data base that was established in 1976 provided easy access to data concerning organic chemicals and their production in a format that facilitated comparisons of various aspects of the industry. It served as a tool for understanding the organic chemical industry, for guiding EPA work in a knowledgeable and systematic manner, and for increasing work effort efficiency.

Since 1977 Radian has updated the data in the OCPDB and increased the size and capabilities of the system. This interim report describes progress made toward this objective. The revised OCPDB was made fully operational in 1979. While this report describes the basic form and substance of the system, it is not meant to imply that the system is static. Changes, expansions, and improvements are expected, as the needs arise. The new system is expected to be even more responsive to

changes in program needs and will allow more flexibility in operation.

### Discussion and Procedure

The original OCPDB consisted of a matrix of about 300 chemicals and their 610 production sites. The chemical list was begun with the one compiled by Monsanto Research Corporation under EPA Contract 68-02-1320. Several additions were made to complete the list. The basic petrochemical feedstocks were added: toluene, xylene, ethylene, propylene, C<sub>2</sub>-C<sub>4</sub> hydrocarbons. Also added were chemicals that had production volumes equal to or greater than those chemicals already included (about 10 million pounds per year). Prioritized lists of toxic chemicals were examined, and chemicals not in the data base were added. The list was then compared to the list generated in the Source Assessment Program (EPA Contract 68-02-1874) to check for omissions. Production sites for the chemicals were obtained from the open literature, and this file formed the other dimension of the computerized matrix. The mechanical structures of the chemical and producer data files are shown in Figures 1 and 2.

New adjectives or descriptors were added to describe chemical entries. These include Chemical Abstracts Services (CAS) registry numbers, process routes, additional toxicity data, use descriptions, sales, and synonyms. New data files describing production sites include parent company name and river basins. The data files from the original OCPDB have been updated in cases in which new data have become available.

Table 1 lists all of the data files included in the OCPDB system. It also shows the number of unique data items, the total number of occurrences of data items within a file, and the amount of computer storage required for each file.

The revised data base has been implemented with a data base management system (DBMS): System 2000®. A DBMS was chosen to eliminate inefficiencies in the original system. System 2000® was selected because it is well-proven and widely used in commercial and governmental institutions, including EPA.

Flexibility in reporting is another major improvement in the revised

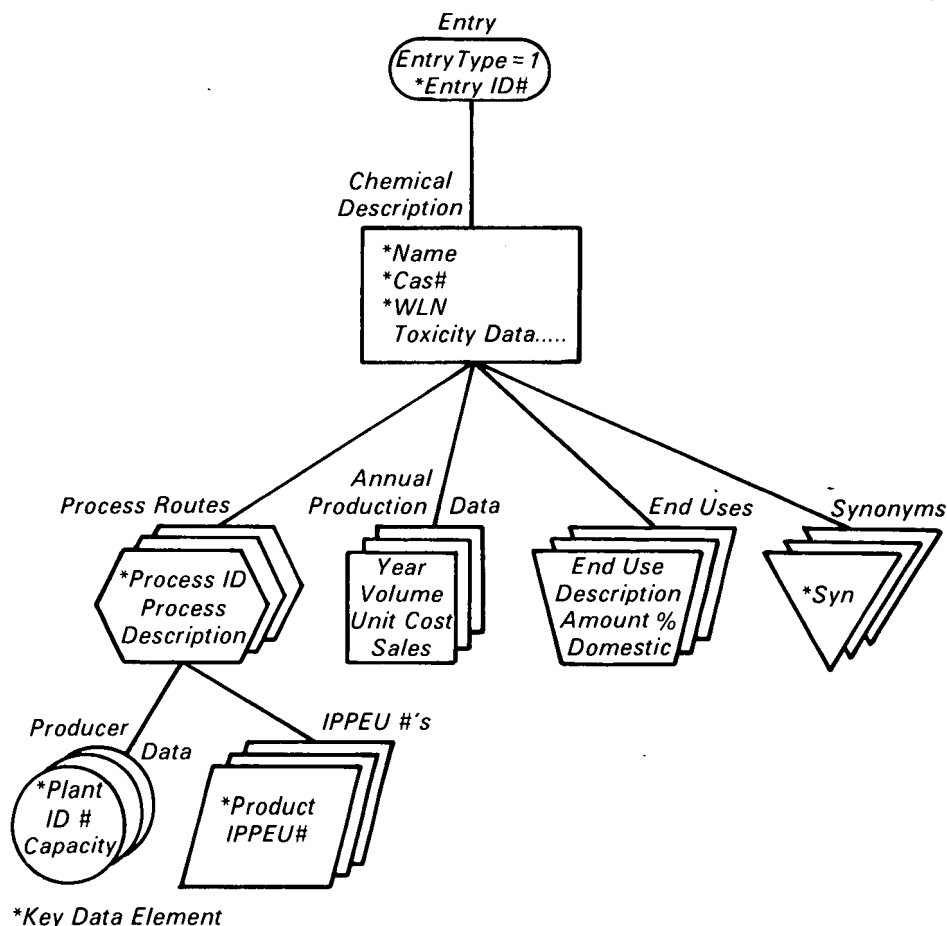


Figure 1. Hierarchical structure of the chemical data files of the OCPDB.

system. Both interactive and batch modes of access are possible, and additional report formats may be defined at any time. Analytical report capabilities such as the "minimum sites search" have been retained and expanded through the possibility of linking the DBMS to other data handling systems.

The sections of the report describe in more detail the system contents and its workings. Section 1 is an introduction to the system and its revision. Section 2 is a summary provided for those who want a quick overview of the OCPDB. A description of the data contained in the data base is presented in Section 3. The system mechanics and structure are detailed in Section 4. The capabilities of the system and an understanding of its uses can best be gained in Section 5 which contains sample reports and example access modes. Several long tables and technical sections have been appended to facilitate reading the report. Appendix A is a listing of OCPDB chemicals; Appendix B is a listing of

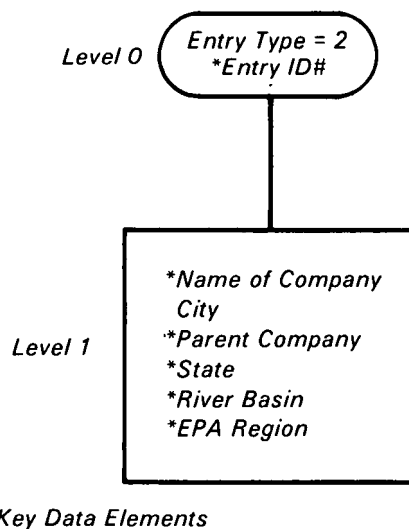


Figure 2. Hierarchical structure of the producer data files of the OCPDB.

System 2000® is a registered trademark of MRI Systems Corporation.

**Table 1. Data Tally for the OCPDB<sup>a</sup>**

	No. of Unique Values	No. of Occurrences	Total Computer Storage Volume in No. of Bytes
Entry Type	2	1,944	1,944
Entry ID	1,621	1,944	8,720
<b>Chemical Related Data</b>			
OCPDB Chemicals	597	597	14,925
CAS Numbers	518	525	5,250
New Chemical Markers	1	224	224
Priority Pollutant Markers	1	135	135
Wiswesser Line Notation	500	506	20,240
Process ID	8	1,131	2,262
Process Description	640	1,133	28,325
IPPEU Numbers	224	318	954
Uses Description	1,794	2,763	138,150
Use Volume	b	784	6,272
Use by % of Consumption	b	711	4,266
Use IPPEU Numbers	90	106	318
Synonyms	5,285	5,426	217,040
<b>Toxicity Data</b>			
NIOSH Registry Number	b	439	3,512
LD <sub>50</sub> Mode	6	335	1,005
LD <sub>50</sub> Species	6	335	1,005
LD <sub>50</sub> Amount	b	335	2,680
LD <sub>50</sub> Units	4	597	5,970
LC <sub>LO</sub> Mode	3	134	402
LC <sub>LO</sub> Species	9	134	402
LC <sub>LO</sub> Amount	b	135	1,080
LC <sub>LO</sub> Units	37	597	5,970
AQTX	b	170	2,496
TLV	b	185	1,850
TLV Units	5	597	5,970
Sax Ratings	b	1,470	1,470
<b>Economic Data</b>			
Year	15	649	1,298
Production Volume	b	287	2,296
Unit Cost	b	348	2,784
Sales	b	176	1,408
<b>Producer Related Data</b>			
Plant ID	1,246	3,703	18,515
Plant Capacity	b	1,190	10,710
Company Names	615	1,346	33,650
Cities	748	1,346	26,920
States	49	1,346	2,692
River Basin	321	1,082	32,460
River Basin Code	326	1,081	8,648
Parent Companies	182	340	8,500
<b>TOTALS</b>	<b>14,853</b>	<b>36,604</b>	<b>632,718<sup>c</sup></b>

<sup>a</sup>As of the date of this interim report.

<sup>b</sup>For non-key data files, the number of unique values is not known.

<sup>c</sup>With indices and other internal data base tables, the total number for the entire data base is approximately 2,000,000 bytes.

OCPDB producers. River basins in which OCPDB producers are located are listed in Appendix C.

Information about accessing the

system may be obtained by contacting the project officer. This publication is a summary of the complete project report, which can be purchased from the

**Table 2. List of Key Data Elements in OCPDB**

<b>Chemical Product Entries</b>	
Name	
OCPDB Number	
CAS Number	
Wiswesser Line Notation	
Use Description	
Original OCPDB Chemical Indicator	
Priority Pollutant Indicator	
Synonym	
Process OCPDB Number	
Process IPPEU Number	
<b>Producer Entries</b>	
OCPDB Number	
Parent Company Name	
Producing Company Name	
City	
State	
River Basin	
EPA Region	

National Technical Information Service.

Three major types of directed retrieval are possible using the OCPDB: information about chemical products, information about producers, and relationships between producers and products. Retrieval is possible through use of any of a number of "key" data elements which are listed in Table 2. Using these retrieval "keys" to access the files, the sorting, filing and reporting possibilities are virtually limitless.

---

*G. E. Wilkins, C. H. Tucker, and E. D. Gibson are with Radian Corporation, Austin, TX 78766.*

***Ms. Audrey McBath** is the EPA Project Officer (see below)*

*The complete report, entitled "The Revised Organic Chemical Producers Data Base System," (Order No. PB 199 805; Cost: \$11.00 subject to change) will be available from:*

*National Technical Information Service*

*5285 Port Royal Road*

*Springfield, VA 22161*

*Telephone: 703-557-4650*

*The EPA Project Officer can be contacted at:*

*Industrial Environmental Research Laboratory*

*U.S. Environmental Protection Agency*

*Cincinnati, OH 45268*