Pesticide Use Report Data

User Guide & Documentation

CD-ROM Media

California Department of Pesticide Regulation

Revised July 2002

Table of Contents

Chapter
1

Using this Documentation

This documentation is designed to assist you with the use of the Pesticide Use Report Data on CD-ROM media. It contains an overview of the Pesticide Use Reporting database and a description of the structure of the Pesticide Use Report Data and respective lookup tables. Data dictionaries provide additional information about the structure, appropriate use, and limitations of each data field. A diagram which depicts the relationships between the pesticide use data and all the lookup tables is located at the end of Chapter 2.

Who to Contact

This documentation package has been provided by the State of California, Department of Pesticide Regulation. If you have any questions, please contact the Pest Management and Licensing Branch at: (916) 324-4100. Our mailing address is P.O. Box 4015, Sacramento, CA 95812-4015.

Overview of the Pesticide Use Report Data

Building The Use Report Data

Each pesticide use record on the CD-ROM is referred to as a Use Data Chemical (UDC) record. The record contains information for an individual active ingredient contained in the product used in an application. Since pesticide applications are reported on a product basis, and since a product may contain multiple chemicals, there may be several UDC records for a single application of one product.

Each pesticide application is assigned a unique identification number at the time it is processed; this number (field: USE_NO) in combination with the chemical code (field: CHEM_CODE), uniquely identifies each of the individual use data records within a given year.

The information in the UDC record comes from several sources: the Use Report Transaction Record (USE), the product table (PRODUCT), the product/site table (PROD_SITE), and the product/chemical table (PROD_CHEM). The use report transaction record contains information submitted by the grower or applicator about an application of pesticide use. On an agricultural application, this includes what product was used, the property operator, where the application was made, the commodity to which the application was made, when the application was made, and how much product was applied. When a use report transaction record is processed, information such as the chemical codes, chemical percent, and product information is retrieved and data is verified using the PRODUCT, SITE, COUNTY and PROD_CHEM tables.

The Lookup Tables

The following tables are used to decode many of the data fields in the UDC and Product tables. A relationship diagram between the tables is provided at the end of Chapter 2.

PRODUCT: Information unique to the product such as name, California Registration Number, product number (prodno), registration status, formulation, etc.

FORMULA: Information regarding the formula composition of products; decodes the FORMULA_CD field.

CHEM_CAS: Chemical Abstract Service (CAS) Number lookup table. The CAS number is a numeric designation that is given to a specific chemical compound by the Chemical Abstract Service. The values for CHEM_CODE are not unique since a chemical may have more than one CAS number.

CHEMICAL: Cross-reference table with the chemical codes and chemical names used by DPR. The values for CHEM_CODE are unique.

SITE: Cross-reference table with site codes and site names (or **commodity name**). DPR considers a crop or commodity upon which chemicals can be used as a 'site' . Each commodity is given a 'SITE_CODE'. This code can be found in the use record and in the SITE lookup table.

COUNTY: Cross-reference table with county codes and county names.

QUALIFY: Cross-reference table of qualifier codes used with commodities to provide more specificity of description.

The California Registration number and PRODNO database field

Each pesticide product is identified by a four-part California Registration Number. The fields making up this number are: MFG_FIRMNO, LABEL_SEQ_NO, REVISION_NO, and REG_FIRMNO. The product registration number usually does not appear on the product label in this format; it may appear only as the first two of the four parts. The first two (MFG_FIRMNO, LABEL_SEQ_NO) are usually assigned by the US EPA. These fields represent the US EPA number for the company (MFG_FIRMNO) and a product sequence within the company (LABEL_SEQ_NO). California appends a revision code (REVISION_NO) to the US EPA registration number to identify alternate brand names with the same mfg_firmno and label_seq_no. An additional field (REG_FIRMNO) designates the firm registering the product. If the registering firm is different than the manufacturing firm, this number will also be on the physical product label. When one company manufactures the product and another registers it as a subordinate seller, it is considered a “sub-registration.” Approximately 25% of the products registered in California are sub-registrations.

Each product in the PRODUCT table is assigned a unique product number (field: PRODNO). The prodno field is used to facilitate data processing and to associate product information maintained in various tables without having to match the four-part California Registration Number. For the purposes of this data set, the prodno field can be used to join information in the UDC and PRODUCT tables.

Each chemical registered as an active ingredient is assigned a unique code number (CHEM_CODE). The portion of the formulated product that is not identified as 'active ingredient' is

consolidated into a single code for 'inert ingredient'. While some chemicals that are listed in formulated products as inert ingredients are of toxicological concern, their identity and percentage in the product is considered 'Confidential Business Information' under the provisions of the Federal Insecticide, Fungicide, and Rodenticide Act, and is not available on these public CD-ROMs.

Identifying Probable Errors: The Outlier Program

To improve data quality, DPR developed a statistical method to detect probable errors in the data fields for acres treated and the pounds of pesticide used. Called the outlier program, this method calculates pesticide use rates (pounds of active ingredient applied divided by acres treated) that are then examined using a variety of statistical methods. The records with highly unlikely use rates (outliers) are identified, thereby serving to flag suspect pesticide use records.

Errors can occur, for example, if a decimal is misplaced, if the measure is incorrect, if the number of acres or units treated is incorrect, or if the diluted amount is reported. We used three different criteria to identify outliers by comparing each use rate with an estimate of the maximum rate for that type of use.

These flags are given in a separate table on the CD-ROM, named **outlyYY.txt**, where YY is the last two digits of the year. Detailed information about the outlier criteria and the structure and use of the outlier table is given in Chapter 4.

The Pesticide Use Report Data on CD-ROM

The Pesticide Use Report Data is comprised of several million records per year. Since this large amount of data is unmanageable for many customers who are interested only in regional subsets of the data, the use records have been separated by county. Each CD-ROM contains 58 data files (one for each county), and the lookup tables. The pesticide use data files begin with the naming convention: **UDCyy_nn**. UDC is for Use Data Chemical, yy is for the year of the data, and nn is used to represent the county code, i.e. UDC99_01 represents Use Data Chemical for 1999 for county 01 (Alameda).

The Pesticide Use Reporting data is placed on CD-ROM in DOS text with comma-delimited fields. The first line provides the field headings.

How To Upload the Data

Due to the large number of records in most of the data files and the limitations in the number of records a spreadsheet application can import, a spreadsheet application cannot be used to manage this data. However, to become familiar with the data structure, a small data file (e.g. udc99_46.txt, the use data for Sierra County) or the first few records of a larger data file can be viewed using a spreadsheet application. Import the data file as comma delimited data. The data files can also be viewed using a word-processing application. To do so, set the "page layout" to letter or legal landscape mode to view the entire length of the record.

In addition to the text CD-ROM, spatial data of the 58 counties is provided on a separate CD-ROM for use with GIS software.

Chapter 2

Record Structures for UDC and Lookup Table data

The following tables define the record structure of each data file. The "Field Seq. No." (field sequence number) identifies the order in which each field appears in the data record and in the data dictionary in Chapters 2 and 3 of this document. The “field name” indicates the name of the field. The “type” indicates whether the field is a Numeric (N), Character (C), or Date (Date) field. “Mask“ displays the field as Numeric, Character, or Date values representing the size of each field along with decimal places (if used).

Use Data Chemical (UDC)

Field Seq. No.	Field Name	Type	Mask
1	USE_NO	N	N(8)
2	PRODNO	N	9999999
3	CHEM_CODE	N	99999
4	PRODCHEM_PCT	N	999.99999
5	LBS_CHM_USED	N	Floating Decimal
6	LBS_PRD_USED	N	N(10).9999
7	AMT_PRD_USED	N	N(8).9999
8	UNIT_OF_MEAS	C	AA
9	ACRE_PLANTED	N	N(8).99
10	UNIT_PLANTED	C	A
11	ACRE_TREATED	N	N(8).99
12	UNIT_TREATED	C	A
13	APPLIC_CNT	N	999999
14	APPLIC_DT	DATE	MMDDYYYY
15	APPLIC_TIME	N	HHMM
16	COUNTY_CD	C	AA
17	BASE_LN_MER	C	A
18	TOWNSHIP	C	AA
19	TSHIP_DIR	C	A
20	RANGE	C	AA
21	RANGE_DIR	C	A
22	SECTION	C	AA
23	SITE_LOC_ID	C	A(8)

Use Data Chemical (UDC) Continued

Field Seq. No.	Field Name		Mask
24	GROWER_ID	C	A(11)
25	LICENSE_NO	C	A(13)
26	PLANTING_SEQ	N	9
27	AER_GND_IND	C	A
28	SITE_CODE	N	999999
29	QUALIFY_CD	N	99
30	BATCH_NO	N	9999
31	DOCUMENT_NO	C	A(8)
32	SUMMARY_CD	N	9999
33	RECORD_ID	C	A

Product Lookup Table

Field Seq. No.	Field Name	Type	Mask
1	PRODNO	N	999999
2	MFG_FIRMNO	N	9999999
3	REG_FIRMNO	N	9999999
4	LABEL_SEQ_NO	N	99999
5	REVISION_NO	C	AA
6	FUT_FIRMNO	N	9999999
7	PRODSTAT_IND	C	A
8	PRODUCT_NAME	C	A(100)
9	SHOW_REGNO	C	A(24)
10	AER_GRND_IND	C	A
11	AGRICCOM_SW	C	A
12	CONFID_SW	C	A
13	DENSITY	N	999.999
14	FORMULA_CD	C	AA
15	FULL_EXP_DT	DATE	MMDDYYYY
16	FULL_ISS_DT	DATE	MMDDYYYY
17	FUMIGANT_SW	C	A
18	GEN_PEST_IND	C	A
19	LASTUP_DT	DATE	MMDDYYYY
20	MFG_REF_SW	C	A
21	PROD_INAC_DT	DATE	MMDDYYYY
22	REG_DT	DATE	MMDDYYYY
23	REG_TYPE_IND	C	A
24	RODENT_SW	C	A
25	SIGNLWRD_IND	N	9
26	SOILAPPL_SW	C	A
27	SPECGRAV_SW	C	A
28	SPEC_GRAVITY	N	99.9999
29	CONDREG_SW	C	A

CAS Number Lookup Table

Field Seq. No.	Field Name	Type	Mask
1	CHEM_CODE	N	99999
2	CAS_NUMBER	C	A(12)

Chemical Code Lookup Table

Field Seq. No.	Field Name	Type	Mask
1	CHEM_CODE	N	99999
2	CHEMALPHA_CD	N	9(8)
3	CHEMNAME	C	A(171)

Site Code Lookup Table

Field Seq. No.	Field Name	Type	Mask
1	SITE_CODE	N	999999
2	SITE_NAME	C	A(50)

Formula Code Lookup Table

Field Seq. No.	Field Name	Type	Mask
1	FORMULA_CD	C	AA
2	FORMULA_DSC	C	A(50)

Qualify Code Lookup Table

Field Seq. No.	Field Name	Type	Mask
1	QUALIFY_CD	N	999
2	QUALIFY_DSC	C	A(50)

County Code Lookup Table

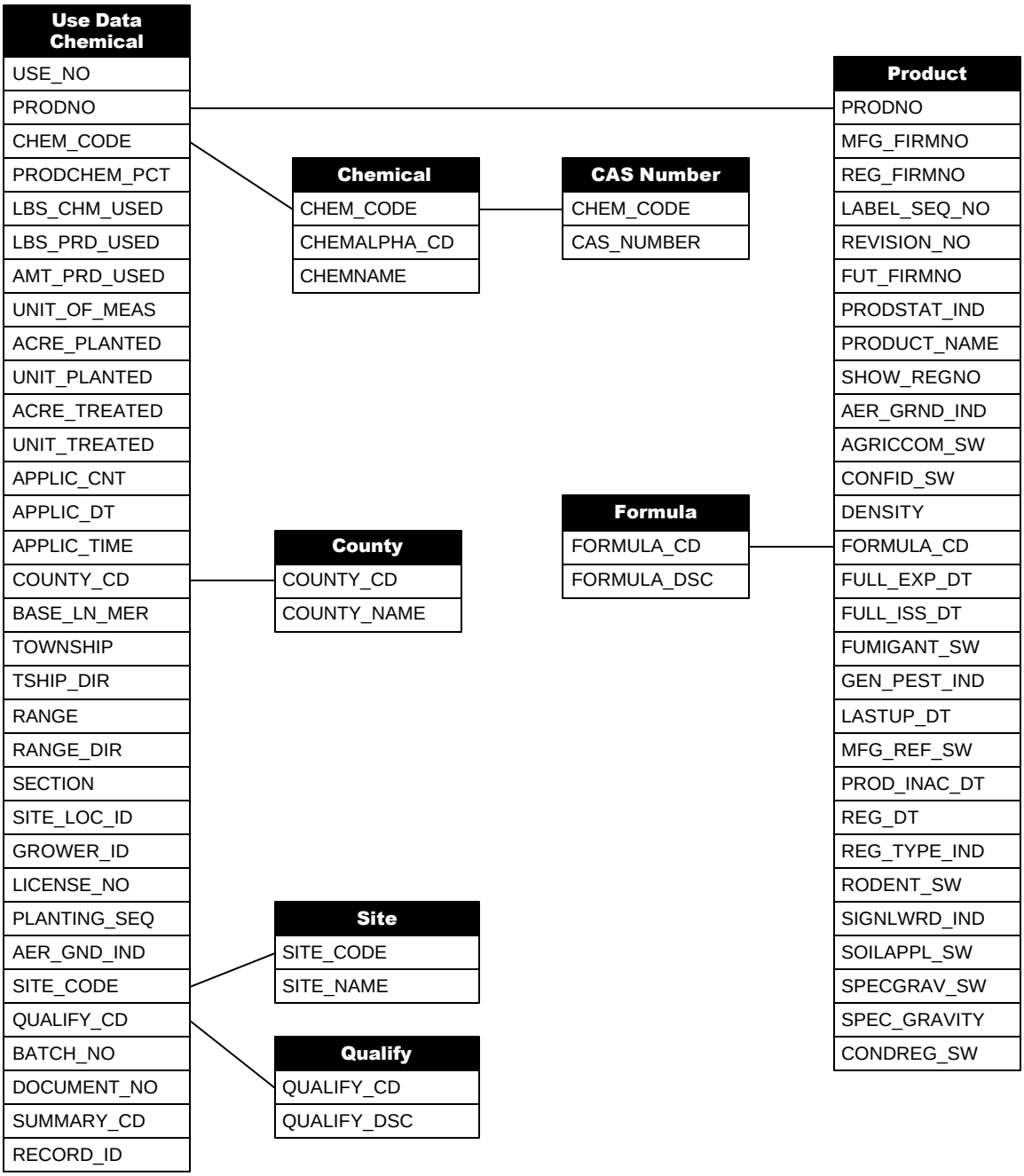
Field Seq. No.	Field Name	Type	Mask
1	COUNTY_CD	C	AA
2	COUNTY_NAME	C	A(15)

County Code Reference List

County Code	County
01	Alameda
02	Alpine
03	Amador
04	Butte
05	Calaveras
06	Colusa
07	Contra Costa
08	Del Norte
09	El Dorado
10	Fresno
11	Glenn
12	Humboldt
13	Imperial
14	Inyo
15	Kern
16	Kings
17	Lake
18	Lassen
19	Los Angeles
20	Madera
21	Marin
22	Mariposa
23	Mendocino
24	Merced
25	Modoc
26	Mono
27	Monterey
28	Napa
29	Nevada

County Code	County
30	Orange
31	Placer
32	Plumas
33	Riverside
34	Sacramento
35	San Benito
36	San Bernardino
37	San Diego
38	San Francisco
39	San Joaquin
40	San Luis Obispo
41	San Mateo
42	Santa Barbara
43	Santa Clara
44	Santa Cruz
45	Shasta
46	Sierra
47	Siskiyou
48	Solano
49	Sonoma
50	Stanislaus
51	Sutter
52	Tehama
53	Trinity
54	Tulare
55	Tuolumne
56	Ventura
57	Yolo
58	Yuba

Pesticide Use Data and Lookup Table Relationship Diagram



Chapter
3

Data Dictionary for the Use Data

How To Use the Data Dictionary

This data dictionary is organized by the Field Sequence Number, which identifies the order in which the field appears in the data record. For example, in a UDC data record, field sequence number 16 is the COUNTY_CD field, representing the county code.

The field name used in the database is given along with the full descriptive field name.

A data type is given to indicate whether the field is Numeric, Character (CHAR), or Date. The format is shown by indicating, for Numeric, the number "9" extended to the actual width of the field; for example, a 7 digit Numeric field would have a format of "9999999" and a 1 digit Numeric field would have a format of "9". A Character field of 7 Characters would have a format of "AAAAAAA" and for longer fields such as a Character field of 100 the format is displayed as A(100).

A description of the field is provided along with important uses or limitations specific to this field.

Notable data validations are also listed.

A blank value for a field may occur. The data may not be required, or the value submitted may have been invalid (for example, a character submitted in a numeric data field) and was replaced with a blank value during data validation.

UDC Table

Field Sequence No. ***1***

Field Name: ***USE_NO ¾*** **Use Number**

Data Type: NUMERIC

Format: 9999999

Description: System assigned sequential number to uniquely identify a pesticide product use record within a year.

Notes: Use this number to identify all chemical records associated with a single application of a product.

Validation: Unique

UDC Table

Field Sequence No. ***2***

Field Name: ***PRODNO ¾*** **Product Number**

Data Type: NUMERIC

Format: 999999

Description: System assigned sequential number to uniquely identify a pesticide product. (Assigned by product label database.) PRODNO has a one-to-one relationship to the four-part key composed of: MFG_FIRMNO + LABEL_SEQ_NO + REVISION_NO + REG_FIRMNO (also known as the California Registration Number). A detailed description of the California Registration Number is given in Chapter 1.

Notes: Reference PRODNO in the Product table.

UDC Table

Field Sequence No.: ***3***

Field Name: ***CHEM_CODE ¾*** **Chemical Code**

Data Type: NUMERIC

Format: 99999

Description: Identifies the active ingredient (AI) contained in the applied product. The CHEM_CODE is sequentially assigned for each new active ingredient during the registration process.

Notes: Assigned to each use_data_chemical record based on the reported California Registration Number (comprised of the MFG_FIRMNO + LABEL_SEQ_NO + REVISION_NO + REG_FIRMNO, i.e. PRODNO in the label database). Applications are reported by product, therefore a separate use_data_chemical record is created for each AI contained in the applied product.

UDC Table

Field Sequence No. ***4***

Field Name: ***PRODCHEM_PCT*** ¾ **Product Chemical Percent**

Data Type: NUMERIC

Format: 999.99999

Description: The percentage of active ingredient found in the product as shown on the product label. This value is extracted from the label database PROD_CHEM table.

Notes:

Validation: Must be a positive number. The total amount of all active and inert ingredients in a product must equal 100%.

UDC Table

Field Sequence No. ***5***

Field Name: ***LBS_CHM_USED ¾*** **Pounds Chemical Used**

Data Type: NUMERIC

Format: floating decimal

Description: Pounds of the active ingredient (AI) in the applied product. The specific AI is identified in the CHEM_CODE field.

Notes: This value is calculated based on the pounds of product used times the percent AI in the product. (LBS_CHM_USED = LBS_PRD_USED x PRODCHEM_PCT/100)

Products with liquid formulations are first converted to pounds of product used and then to pounds of AI used. (See field sequence 6.)

Pounds of chemical used is used to calculate the total pounds of AI in the annual *Summary of Pesticide Use Report*.

UDC Table

Field Sequence No. ***6***

Field Name: ***LBS_PRD_USED*** ¾ **Pounds Product Used**

Data Type: NUMERIC

Format: N(10).9999

Description: Pounds of product applied. The pounds of product applied can include products applied to soil, commodities, etc.

Notes: Pounds product applied for each application is calculated based on the product's formulation (liquid or dry), specific gravity (if applicable), amount of product reported used (AMT_PRD_USED), and unit of measure (UNIT_OF_MEAS).

This field is used to derive pounds chemical used as stated in the field sequence 5.

UDC Table

Field Sequence No. ***7***

Field Name: ***AMT_PRD_USED ¾*** **Amount Product Used**

Data Type: NUMERIC

Format: N(8).9999

Description: Amount of product reported used. This value is converted to pounds of product used (LBS_PRD_USED) during the data loading process.

Notes: See UNIT_OF_MEAS (field sequence number 8) for related units.

UDC Table

Field Sequence No. ***8***

Field Name: ***UNIT_OF_MEAS ¾*** **Unit of Measure**

Data Type: CHAR

Format: AA

Description: Refers to the unit of measure in conjunction with the reported AMT_PRD_USED field (field sequence no. 7). It is used to convert units applied to the common unit of pounds.

Notes: Codes are:

OZ	ounces
LB	pounds
PT	pints
QT	quarts
GA	gallons
ML	milliliters
LI	liters
GR	grams
KG	kilograms

Validation: OZ, LB, PT, QT, GA, ML, LI, GR, KG must be acceptable with the product's formulation type (liquid or dry).

UDC Table

FIELD Sequence No. ***9***

Field Name: ***ACRE_PLANTED*** ¾ **Acres Planted**

Data Type: NUMERIC

Format: N(8).99

Description: Size of field, or other unit (e.g. number of tree trunks), which was planted with an agricultural commodity.

Notes: See UNIT_PLANTED (field sequence no. 10) for related units.

For example, if the application occurred on a planted field of 100 acres, then ACRE_PLANTED = 100, and UNIT_PLANTED = A.

When an application is reported in units of square feet, it is converted to acres for internal validation processes.

This field cannot be used to calculate agricultural statistics such as the total amount (acres) planted of a particular commodity because there could be multiple applications to the same field (site ID) and commodity (site) combination during a year.

Acres planted on the same site_id may change during a year due to multiple plantings.

Validation: Must be greater than or equal to ACRE_TREATED. Must be a numeric value. Must be filled when using site (commodity) codes greater than 100.

Prior to 7/95, validation for nursery codes (151 - 156) was not performed.

UDC Table

Field Sequence No. ***10***

Field Name: ***UNIT_PLANTED ¾*** **Unit Type (Planted)**

Data Type: CHAR

Format: A

Description: Refers to the type of units planted in conjunction with the reported ACRE_PLANTED field.

Codes are:

A	acres
S	square feet
C	cubic feet
K	thousand cubic feet
U	Misc. Examples of misc. units include: bins, tree holes, bunches, pallets, etc.

Prior to 7/95, validation for nursery codes (151 - 156) was not performed.

Validation: A, S, C, K, and U. UNIT_PLANTED must be consistent with UNIT_TREATED (field sequence no. 12).

UDC Table

Field Sequence No. ***11***

Field Name: ***ACRE_TREATED*** ¾ **Acres Treated**

Data Type: NUMERIC

Format: N(8).99

Description: Number of units upon which a pesticide product was applied. The name of this field is misleading since the value in this field does not necessarily represent acreage. See UNIT_TREATED field (field sequence no. 12) for related units.

Notes: For example, an application to 100 chicken houses would be reported as ACRE_TREATED = 100, with UNIT_TREATED = U (miscellaneous).

If the application occurred on a planted field of 100 acres, then ACRE_TREATED = 100, and UNIT_TREATED = A. If the application occurred on only 50 acres of a 100-acre field, then ACRE_TREATED = 50.

Applications reported in square feet are converted to acres for internal validation process.

The location of actual ground areas receiving the application within a field cannot be distinguished. For example: two 50-acre applications within a 100-acre field may have occurred twice on the same ground, or may represent single applications to two distinct 50 acre subdivisions.

Validation: Must be less than or equal to ACRE_PLANTED. Must be a numeric value.

Prior to 7/95, validation for nursery codes (151 - 156) was not performed.

UDC Table

Field Sequence No.: ***12***

Field Name: ***UNIT_TREATED*** ¾ **Unit Type (Treated)**

Data Type: CHAR

Format: A

Description: Refers to the type of units treated in conjunction with the reported ACRE_TREATED field.

Notes: Possible values are:

A	acres
S	square feet
C	cubic feet
K	thousand cubic feet)
P	pounds
T	tons
U	misc. unit (Examples of misc. units (U) include: bins, tree holes, bunches, pallets, etc.)

Prior to 7/95, validation was performed only for site codes equal to or greater than 1000.

Validation: A, S, C, K, P, T, and U. UNIT_ TREATED must be consistent with UNIT_ PLANTED (field sequence no. 10).

UDC Table

Field Sequence No. ***13***

Field Name: ***APPLIC_CNT ¾*** **Application Count**

Data Type: NUMERIC

Format: 999999

Description: Total number of applications for each product used by an operator performed during the reporting month as noted on a non-production monthly summary report.

Notes: For production agriculture applications, the system inserts a one. For non-production applications, the application count may or may not be reported.

Validation: Must be a numeric value.

UDC Table

Field Sequence No. ***14***

Field Name: ***APPLIC_DT ¾*** **Application Date**

Data Type: DATE

Format: MMDDYYYY

Description: Date that the pesticide product was applied.

Notes: The application date used for a non-production summary records (RECORD_ID = '2' or 'C') does not reflect the actual date of application. (DD = 01 except for CalTrans records where DD = 28)

Validation: Must be prior to the current system date (the date the use report data is loaded and verified), and the year must be consistent with current use reporting year. Must be a valid date.

UDC Table

Field Sequence No. ***15***

Field Name: ***APPLIC_TIME*** ¾ **Application Time**

Data Type: NUMERIC

Format: HHMM

Description: Time that the pesticide product application was completed.

Notes: The application time is military format and is only included on production agricultural reports, record types A, B, 1, and 4.

This field was added to the database in 1999 but was inconsistently downloaded until 2000.

Validation: Must be a valid time.

Required for record types A, B, 1, and 4.

UDC Table

Field Sequence No. ***16***

Field Name: ***COUNTY_CD ¾*** **County Code**

Data Type: CHAR

Format: AA

Description: County code established by numbering an alphabetized list of California's 58 counties.
For example, '01' = Alameda; '58' = Yuba.

Notes: See “County” table on CD-ROM and “County Code Reference List” in Chapter 2.

Validation: Must be 01 - 58. Not blank.

UDC Table

Field Sequence No. ***17***

Field Name: ***BASE_LN_MER*** ¾ **Base Line & Meridian**

Data Type: CHAR

Format: A

Description: Public Lands Survey (PLS) System Base Line & Meridian for the application location. Often referred to as the 'Meridian'.

Notes: The Base Line and Meridian establish a point of reference for determining locations using the PLS system. There are three Base/Meridians in California. Combination of the county, meridian, township, range and section fields identifies a unique location within the PLS.

Validation: Codes for the meridians in California are: S (San Bernardino), M (Mount Diablo), H (Humboldt).

UDC Table

Field Sequence No. ***18***

Field Name: ***TOWNSHIP ¾*** **Township Number**

Data Type: CHAR

Format: AA

Description: Number of the township in the Public Land Survey System where the application occurred. Must be combined with BASE_LN_MER and TSHIP_DIR to determine the unique township.

Notes: A township will not appear, or will be '00' on use reports where location information is not required (e.g. structural, landscape. etc). Each meridian, township, and range combination on a use report must fall within the reported county.

Validation: 01-48

Required for record types A, B, E, F, 1, and 4.

UDC Table

Field Sequence No. ***19***

Field Name: ***TSHIP_DIR ¾*** **Township Direction**

Data Type: CHAR

Format: A

Description: Public Land Survey System direction from a base line. Townships are numbered to the north and south from an east/west-running base line.

Notes:

Validation: Must be N (North) or S (South).

Required for record types A, B, E, F, 1, and 4.

UDC Table

Field Sequence No. ***20***

Field name: ***RANGE ¾ Range***

Data Type: CHAR

Format: AA

Description: Number of the range within the Public Land Survey System where the application occurred. Must be combined with BASE_LN_MER and RANGE_DIR to determine the unique range.

Notes: A range will not appear, or will be '00' on use reports where location information is not required (e.g. structural, landscape. etc). Each meridian, township, and range combination on a use report must fall within the reported county.

Validation: 01 to 47

Required for record types A, B, E, F, 1, and 4.

UDC Table

Field Sequence No. ***21***

Field Name: ***RANGE_DIR*** ¾ **Range Direction**

Data Type: CHAR

Format: A

Description: Public Land Survey System direction for the range where an application was reported. Ranges are numbered to the east and west from a north/south-running base meridian.

Notes:

Validation: Valid values are: E (East), W (West).

Required for record types A, B, E, F, 1, and 4.

UDC Table

Field Sequence No. ***22***

Field Name: ***SECTION*** ***¾*** **Section**

Data Type: CHAR

Format: AA

Description: An area of approximately one square mile (640 acres) within the Public Land Survey System where the pesticide application occurred.

Each township may be divided into a maximum of 36 sections. Must be used in combination with meridian, township, and range to identify the unique section.

Notes:

Validation: 01 to 36.

Required for record types A, B, E, F, 1, and 4.

UDC Table

Field Sequence No. ***23***

Field Name: ***SITE_LOC_ID ¾*** **Site Location ID**

Data Type: CHAR

Format: AAAAAAAA

Description: Also known as **Site ID**. A code assigned by the County Agricultural Commissioner (CAC) on the use permit which indicates a particular location (field) where an application may occur.

Notes: It was designed to uniquely identify geographic field locations, but is currently assigned at the discretion of individual CACs and growers.

Validation:

UDC Table

Field Sequence No. ***24***

Field Name: ***GROWER_ID ¾*** **Grower Identification Number**

Data Type: CHAR

Format: AAAAAAAAAAA

Description: Number assigned to a grower or property operator by the County Agricultural Commissioner. Also known as the **permit number**, or **operator identification number**.

Notes: The number is composed of:

reporting county_cd	two digits representing the county where the pesticide application occurred.
application year	the last two digits of the year when the application occurred
home county_cd	two digits representing the county where the owner/operator resides
"permit" number	an arbitrary five-digit number assigned to the owner/operator

The last seven digits of the grower_id may be used to identify individual owner/operator. DPR does not collect names and addresses; that information is only available from the County Agricultural Commissioner. The GROWER_ID + SITE_LOC_ID may identify a unique agricultural parcel or field.

Validation: County codes must be 01 – 58, and the year must be consistent with the year of application.

UDC Table

Field Sequence No. ***25***

Field Name: ***LICENSE_NO ¾*** **License Number**

Data Type: CHAR

Format: A(13)

Description: PCO license number.

Notes: This field was added to the database in 1999 but was inconsistently downloaded until 2000.

Validation: This field is required for type ‘C’ and ‘2’ records only. It is required if the GROWER_ID field is left blank, otherwise optional.

UDC Table

Field Sequence No. ***26***

Field Name: ***PLANTING_SEQ*** ¾ **Planting Sequence**

Data Type: NUMERIC

Format: 9

Description: Number to indicate multiple plantings of the same crop or commodity at the same SITE_LOC_ID (site location identification). Not uniformly used; not validated.

Notes:

Validation: None

UDC Table

Field Sequence No. ***27***

Field Name: ***AER_GND_IND*** ¾ **Aerial/Ground Indicator**

Data Type: CHAR

Format: A

Description: Also known as the **Air/Ground Application Flag,** or **Method of Application**. Indicates whether the product was applied by air, ground, or other equipment.

A Aerially applied
G Ground (ground-based equipment) applied
O Other application methods

Notes: Other application methods (O) may include: paint, ear tag, dip, injection, chemigation, etc.

Validation: Must be A, G, or O.

UDC Table

Field Sequence No. ***28***

Field Name: ***SITE_CODE ¾*** **Site Code**

Data Type: NUMERIC

Format: 999999

Description: Site code from a list established by USEPA and modified for use by DPR. Indicates the target site to which a pesticide product was applied. Also known as **Commodity Code**.

Notes: This code does not refer to the spatial location of a pesticide application (e.g. a field).

Validation: Must be a valid code from the list approved for pesticide use reporting. See site code table.

UDC Table

Field Sequence No. ***29***

Field Name: ***QUALIFY_CD ¾*** **Qualifier Code**

Data Type: NUMERIC

Format: 99

Description: The qualifier code modifies or limits the meaning of the site code upon which the product was applied. Example: '04' indicates 'grown for seed'.

Notes: Some counties use this code to keep more detailed records of crop types or varieties, but it is not a required field. See qualify table for codes and descriptions.

Validation:

UDC Table

Field Sequence No. ***30***

Field Name: ***BATCH_NO ¾*** **Batch Number**

Data Type: NUMERIC

Format: 9999

Description: Sequential number assigned to a file during the download process. This field is useful during trouble shooting and error correction investigations.

Notes: Used as part of inventory control for manual key data entry, or for transmitting data from counties to DPR. This field is not included in the 1990 PUR database. The number is not unique.

Validation:

UDC Table

Field Sequence No. ***31***

Field Name: ***DOCUMENT_NO ¾*** **Document Number**

Data Type: CHAR

Format: AAAAAAAA

Description: Internal sequential tracking number (non-unique) assigned at the time of data entry. Within DPR, the document number refers to a physical piece of paper within a batch of use reports.

Notes: This field is part of a document and line item identifier for physical inventory control. For the internal DPR data entry process, it is combined with process month, batch number and document sequence number (a.k.a. summary code) to uniquely identify an input record. May be used differently by individual counties for their own internal tracking systems.

Validation:

UDC Table

Field Sequence No. ***32***

Field Name: ***SUMMARY_CD ¾*** **Summary Code**

Data Type: NUMERIC

Format: 9999

Description: The line number found within the document for most record types. Indicates how many “lines” (records) are contained on a hard copy use report. For files with over 9,999 records, the far right digit is dropped. (The original database structure was set up for lines per report, not lines per file.)

Notes: For internal use only.

Validation:

UDC Table

Field Sequence No. ***33***

Field Name: ***RECORD_ID ¾*** **Record Identification Number**

Data Type: CHAR

Format: A

Description: Identifies the agency that input a use record, and whether the record is for an individual application or is summarized data. Input agencies are DPR, county agricultural commissioner's (CAC) office, and Prison Industries Authority (PIA).

Notes: Codes for record type and input agency are:

Record Type	DPR	CAC	PIA
Daily Production Ag	1	A	E
Monthly Production Ag	4	B	F
Non-production Summary	2	C	G

Validation:

Chapter
3

Data Dictionary for the Lookup Tables

How To Use the Data Dictionary

This data dictionary is organized by the Table Name and Field Sequence Number which identifies the order in which the field appears in the data record. The table name appears at the top of each page. The field sequence no. identifies the order in which each field appears in the data record and in the data dictionary.

The field name used in the database is given along with the full descriptive field name.

A data type is given to indicate whether the field is Numeric, Character (CHAR), or Date. The format is shown by indicating, for Numeric, the number "9" extended to the actual width of the field; for example, a 7 digit Numeric field would have a format of "9999999" and a 1 digit Numeric field would have a format of "9". A Character field of 7 Characters would have a format of "AAAAAAA" and for longer fields such as a Character field of 100 the format is displayed as A(100).

A description of the field is provided along with important uses or limitations specific to this field.

Notable data validations are also listed.

Product Table

Field Sequence No. ***1***

Field Name: ***PRODNO ¾*** **Product Number**

Data Type: NUMERIC

Format: 999999

Description: System sequentially assigned product number used internally in the database. PRODNO has a one-to-one relationship to the four-part key composed of: MFG_FIRMNO + LABEL_SEQ_NO + REVISION_NO + REG_FIRMNO (AKA California Registration Number).

Notes:

Validation: Must be found on the master label file.

Product Table

Field Sequence No. ***2***

Field Name: ***MFG_FIRMNO*** ¾ **Manufacturer Firm Number**

Data Type: NUMERIC

Format: 9999999

Description: Numeric code assigned by the USEPA to the manufacturing company. Assigned by California if the company has no products registered with US EPA (i.e. is a 'California only' registration). One of the four parts of the California Registration Number.

Notes:

Validation: Must be numeric.

Product Table

Field Sequence No. ***3***

Field Name: ***REG_FIRMNO ¾*** **Registration Firm Number**

Data Type: NUMERIC

Format: 9999999

Description: Numeric code assigned by the USEPA to the registrant firm. (One of four parts of the California Registration Number) Also known as the **sub-registration number** if it is different than the manufacturing firm number.

Notes: If this field is empty, it implies that the manufacturer is also the registrant.

Validation: Must be numeric.

Product Table

Field Sequence No. ***4***

FIELD NAME: ***LABEL_SEQ_NO ¾ Label Sequence Number***

Data Type: NUMERIC

Format: 99999

Description: Sequence number assigned by the USEPA for a new product within the manufacturer company. (One of four parts of the California Registration Number)

Notes: USEPA registration number consists only of the manufacturer firm number, label sequence number, and the registrant firm number (if different than the manufacturer firm number).

Validation: Must be numeric.

Product Table

Field Sequence No. ***5***

Field Name ***REVISION_NO ¾*** **Revision Number**

Data Type: CHAR

Format: AA

Description: The revision code (or alpha code). (One of four parts of the California Registration Number)

Notes: Values 'AA' through 'ZZ'.

If not reported on the use report, this value defaults to 'AA'.

The field is used to validate whether a product exists, but is not used to validate use of the product on the reported commodity (site).

Products whose registration numbers vary only by the revision code have 'no substantive change' to product formulation. (In addition, products that are "sub-registered" have no substantive change in product formulation compared to the product that was originally registered.)

Validation:

Product Table

Field Sequence No. ***6***

Field Name: ***FUT_FIRMNO ¾*** **Future Firm Number**

Data Type: NUMERIC

Format: 9999999

Description: This field was included for future use should new relationships need to be identified

Notes: Reserved for future use.

Validation:

Product Table

Field Sequence No. ***7***

Field Name: ***PRODSTAT_IND ¾*** **Product Status Indicator**

Data Type: CHAR

Format: A

Description: Indicates the product registration status (e.g. currently registered, suspended, etc.)

Notes: May be used to screen active versus inactive products.

Codes are:

A	Active
B	Inactive
C	Inactive, Not Renewed
D	Inactive, Voluntary Cancellation
E	Inactive, Cancellation
F	Inactive, Suspended
G	Inactive, Invalid Data
H	Active, Suspended

Validation: A to H

Product Table

Field Sequence No. ***8***

Field Name ***PRODUCT_NAME ¾*** **Product Name**

Data Type: CHAR

Format: A(100)

Description: The name of the product taken from the registered product label. May be modified by DPR's Registration Branch to ensure uniqueness.

Notes:

Validation:

Product Table

Field Sequence No. ***9***

Field Name: ***SHOW_REGNO ¾*** **Show Registration Number**

Data Type: CHAR

Format: A(24)

Description: The California Registration Number of the product formatted for display purposes.

Notes: This field is up to 24 characters in length. (xxxxxxx-xxxxx-xx-xxxxxxx)

Validation:

Product Table

Field Sequence No. ***10***

Field Name: ***AER_GRND_IND*** ¾ **Aerial/Ground Applicator Indicator**

Data Type: CHAR

Format: A

Description: Air/Ground Application Flag, or Method of Application. Indicates by which method a product can be applied.

Codes are:
A Aerial (applied by airplane)
B Ground-based equipment (**not** applied by airplane)
C Aerial / Ground (can be applied by airplane)

Notes: The product label indicates how the product can be applied.

Validation:

Product Table

Field Sequence No. ***11***

Field Name: ***AGRICCOM_SW ¾*** **Agricultural Commissioner Switch**

Data Type: CHAR

Format: A

Description: The Agricultural Commissioner Use Flag. The flag indicates if the registration is exempt from Mill Assessments. (For internal use only.)

Notes: Extracted from the label database

Validation: Blank (Not Exempt), 'X' (Exempt)

Product Table

Field Sequence No. ***12***

Field Name: ***CONFID_SW ¾*** **Confidential Data Indicator**

Data Type: Char

Format: A

Description: Used to flag adjuvant products which are California-only registered. Chemical formulations in these products are considered confidential.

X = confidential; null = not tracked as confidential

Notes: These products generally can be used on any agricultural commodity. The product/commodity combination is not verified on use reports where this flag is “X”.

Validation:

Product Table

Field Sequence No. ***13***

Field Name: ***DENSITY ¾*** **Density**

Data Type: NUMERIC

Format: 999.999

Description: The weight per unit volume expressed as grams per cubic centimeter for solids and liquids and usually as grams per liter for gasses. The pesticide formula density is in pounds per gallon. Density is derived from the specific gravity at given conditions.

Notes:

Validation:

Product Table

Field Sequence No. ***14***

Field Name: ***FORMULA_CD ¾*** **Formulation Code**

Data Type: CHAR

Format: AA

Description: Formulation of the product. For example: granular, pressurized liquid, emulsifiable concentrate, etc.

Notes: Extracted from the label database. See “Formula” table for codes and descriptions.

Validation:

Product Table

Field Sequence No. ***15***

Field Name: ***FULL_EXP_DT ¾*** **Full Expiration Date**

Data Type: DATE

Format: MMDDYYYY

Description: Expiration date for full product Section 5 (experimental use) or Section 18 (emergency exemption) registrations.

Notes:

Validation:

Product Table

Field Sequence No. ***16***

Field Name: ***FULL_ISS_DT*** ¾ **Full Issuance Date**

Data Type: Date

Format: MMDDYYYY

Description: The issue date of full product Section 5 (experimental use), Section 18 (emergency exemption), or Section 24c (special local need) registrations.

Notes:

Validation:

Product Table

Field Sequence No. ***17***

Field Name: ***FUMIGANT_SW ¾*** **Fumigant Flag**

Data Type: CHAR

Format: A

Description: Used to indicate that the product is a soil fumigant.

Notes: There is no validation for product /site (commodity) because the product may be used for soil fumigation prior to planting a commodity or no specific commodity is listed on the product label.

Validation: X = soil fumigant

Product Table

Field Sequence No. ***18***

Field Name: ***GEN_PEST_IN ¾*** **General Pesticide Type Indicator**

Data Type: CHAR

Format: A

Description: Indicates the general pesticide type of the **product**.

Codes are:
C Chemical (e.g. sulfur dust)
M Microbial (e.g. Bacillus thurengiensis)
K Both chemical and microbial (e.g. Bacillus thurengiensis sulfur dust)

Notes: Extracted from the label database.

Validation: C, M, or K.

Product Table

Field Sequence No. ***19***

Field Name: ***LASTUP_DT*** ¾ **Last Update Date**

Data Type: DATE

Format: MMDDYYYY

Description: This is the date when a product was last modified. (The date is system generated.)

Notes:

Validation:

Product Table

Field Sequence No. ***20***

Field Name: ***MFG_REF_SW ¾ Manufacturing Reformulation Flag***

Data Type: CHAR

Format: A

Description: Indicates that the product is to be used in the manufacturing, reformulation, or repackaging of other products. There is no end use and the product is not subject to mill assessment.

Notes: Sometimes referred to as “manufacturing use only” product. Extracted from label database.

Validation: ‘X’ or ‘ ‘ (blank).

Product Table

Field Sequence No. ***21***

Field Name: ***PROD_INAC_DT ¾ Product Inactivation Date***

Data Type: DATE

Format: MMDDYYYY

Description: The date when the product became inactive.

Notes: Extracted from the label database.

Validation:

Product Table

Field Sequence No. ***22***

Field Name: ***REG_DT ¾*** **Registration Date**

Data Type: DATE

Format: MMDDYYYY

Description: The date when the product was originally registered with the Department of Pesticide Regulation.

Notes: Extracted from the label database.

Validation:

Product Table

Field Sequence No. ***23***

Field Name: ***REG_TYPE_IND ¾*** **Registration Type Indicator**

Data Type: CHAR

Format: A

Description: The product registration type.
Example: Section 3, Section 18, Section 24c, etc.

Notes: Extracted from the label database.
Codes are:
A Section 3 Regular Registration
B Section 24(c) Full Product SLN
C Section 5 Full Product Federal Experimental Use
D California Registration Only
E Section 18 Full Product Emergency Exemption

Validation:

Product Table

Field Sequence No. ***24***

Field Name: ***RODENT_SW ¾*** **Rodenticide Flag**

Data Type: CHAR

Format: A

Description: Indicates that the product is registered as a rodenticide.

Notes: Used to bypass the product/site (commodity) PUR validation process. Products registered as rodenticides may be reported in conjunction with any site (commodity). The site is incidental to the intended use (rodent control).

Validation: 'X' (yes) or (blank = no)

Product Table

Field Sequence No. ***25***

Field Name: ***SIGNLWRD_IND ¾*** **Signal Word Indicator**

Data Type: NUMERIC

Format: 9

Description: The signal word is printed on the front of the product label and must be one of the following:

1 Danger (Poison)
2 Danger (Only)
3 Warning
4 Caution
5 None

Notes: Labels submitted after January 1995 must show a signal word of Danger (Poison), Danger (Only), Warning or Caution.

Validation:

Product Table

Field Sequence No. ***26***

Field Name: ***SOILAPPL_SW ¾*** **Soil Application Flag**

Data Type: CHAR

Format: A

Description: Indicates if the product can be applied directly to the soil.

Notes: Determination taken from language on the label.

Validation: 'X' (yes) or null (no)

Product Table

Field Sequence No. ***27***

Field Name: ***SPECGRAV_SW ¾*** **Specific Gravity Switch**

Data Type: CHAR

Format: A

Description: Indicates whether the specific gravity is noted on the product application or estimated.

Notes: An estimated value can be calculated by using the density or can default to “1” if unknown.

Validation: A = actual; E = estimated

Product Table

Field Sequence No. ***28***

Field Name: ***SPEC_GRAVITY ¾*** **Specific Gravity**

Data Type: NUMERIC (6,4)

Format: 99.9999

Description: The specific gravity of the product.

Notes: The ratio of the density of a substance to the density of a reference substance; it has no units. For solids and liquids, the reference material is distilled water and for gasses the reference material is air or hydrogen. In both cases the reference material is at a standard temperature and pressure. Some products, such as powders, do not have specific gravity associated with them. The default value is negative one (-1). The specific gravity should be noted on the product application for liquid formulations.

Validation:

Product Table

Field Sequence No. ***29***

Field Name: ***CONDREG_SW — Conditional Registration Switch***

Data Type: CHAR

Format: A

Description: Used for products which are registered but still need to meet certain “conditions.” When all conditions are met, the “switch” is removed.

Notes:

Validation:

Chemical CAS Number Table

Field Sequence No. ***1***

Field Name: ***CHEM_CODE*** ¾ **Chemical Code**

Data Type: NUMERIC

Format: 99999

Description: Identifies the active ingredient (AI) contained in a product.

Notes: Assigned to each use_data_chemical record based on the reported California Registration Number (comprised of the MFG_FIRMNO + LABEL_SEQ_NO + REVISION_NO + REG_FIRMNO, i.e. PRODNO in the label database). Applications are reported by product, therefore a separate use_data_chemical record is created for each AI contained in the applied product.

Validation: CHEMICAL table. Must be a valid chemical code found within DPR's master "Chemical" table.

Chem_cas Table

Field Sequence No. ***2***

Field Name: ***CAS_NUMBER*** ¾ **Chemical Abstract Service Number**

Data Type: CHARACTER

Format: A(12)

Description: The number assigned by the Chemical Abstract Service (CAS) to identify specific chemical compounds.

Notes: A chemical may have more than one CAS number. Not all chemicals have an assigned CAS number.

Validation:

Chemical Table

Field Sequence No. ***1***

Field Name: ***CHEM_CODE*** ¾ **Chemical Code**

Data Type: NUMERIC

Format: 99999

Description: Identifies the active ingredient (AI) contained in the product.

Notes: Assigned to each use_data_chemical record based on the reported California Registration Number (comprised of the MFG_FIRMNO + LABEL_SEQ_NO + REVISION_NO + REG_FIRMNO, i.e. PRODNO in the label database). Applications are reported by product, therefore a separate use_data_chemical record is created for each AI contained in the applied product.

Validation: CHEMICAL table. Must be a valid chemical code found within DPR's master “Chemical” table.

Chemical Table

Field Sequence No. ***2***

Field Name: ***CHEMALPHA_CD*** ¾ **Chemical Alpha Sort Code**

Data Type: NUMERIC

Format: 9999999

Description: Used to sort the chemical names in this table alphabetically.

Notes:

Validation:

Chemical Table

Field Sequence No. ***3***

Field Name: ***CHEMNAME*** ¾ **Chemical Name**

Data Type: CHAR

Format: A(170)

Description: The common chemical name for each active ingredient. Usually as listed on the product label.

Notes: Common names can vary depending on the source and naming conventions.

Validation:

Site Table

Field Sequence No. ***1***

Field Name: ***SITE_CODE*** ¾ **Site Code**

Data Type: NUMERIC

Format: 999999

Description: Site code from a list established by USEPA and modified for use by DPR. Also known as **Commodity Code**.

Notes: Applications may be made on a site, around a site, or to soil prior to planting. This code does not refer to the spatial location of a pesticide application (e.g. a field).

Validation:

Site Table

Field Sequence No. ***2***

Field Name: ***SITE_NAME*** ¾ **Site Name**

Data Type: CHAR

Format: A(50)

Description: Identifies the name of the site (i.e. strawberries, wine grapes, tomatoes, etc.). Also known as **commodity name**.

Notes:

Validation:

Formula Table

Field Sequence No. ***1***

Field Name: ***FORMULA_CD*** ¾ **Formulation Code**

Data Type: CHAR

Format: AA

Description: Encodes a general description of the product formulation. Codes are:

A0	DUST/POWDER
B0	EMULSIFIABLE CONCENTRATE
C0	FLOWABLE CONCENTRATE
D0	GEL, PASTE, CREAM
E0	GRANULAR/FLAKE
F0	IMPREGNATED MATERIAL
G0	MICROENCAPSULATED
H0	OIL
I0	PAINT/COATINGS
J0	PELLET/TABLET/CAKE/BRIQUET
K0	PRESSURIZED DUST
L0	PRESSURIZED GAS
M0	PRESSURIZED LIQUID/SPRAYS/FOGGERS
N0	SOLUBLE POWDER
O0	SOLUTION/LIQUID (READY-TO-USE)
P0	WETTABLE POWDER
Q0	SUSPENSION
R0	DRY FLOWABLE
S0	AQUEOUS CONCENTRATE
T0	OTHER (LIQUID)
U0	OTHER (DRY)

Notes: Extracted from the label database.

Validation:

Formula Table

Field Sequence No. ***2***

Field Name: ***FORMULA_DSC*** ¾ **Formulation Description**

Data Type: CHAR

Format: A(40)

Description: The general description of the product formulation.

Notes:

Validation:

County Table

Field Sequence No. ***1***

Field Name: ***COUNTY_CD*** ¾ **County Code**

Data Type: CHAR

Format: AA

Description: County code established by numbering an alphabetized list of California's 58 counties.
For example, '01' = Alameda; '58' = Yuba.

Notes: See “County” table on CD-ROM and “County Code Reference List” in Chapter 2.

Validation: Must be 01 - 58. Not blank.

County Table

Field Sequence No. ***2***

Field Name: ***COUNTY_NAME*** ¾ **County Name**

Data Type: CHAR

Format: AAAAAAAAAAAAAAA

Description: The county name field translates the county_cd field.

Notes:

Validation: Must be a valid name of one of the 58 counties in California.

Qualify Table

Field Sequence No. ***1***

Field Name: ***QUALIFY_CD*** ¾ **Qualify Code**

Data Type: NUMERIC

Format: 999

Description: The qualifier code modifies or limits the meaning of the site code. For example, '04' means 'grown for seed'.

Notes: If a commodity has more than one qualifier, each SITE_CD/QUALIFY_CD combination will be listed separately in the label database.

Validation:

Qualify Table

Field Sequence No. ***2***

Field Name: ***QUALIFY_DSC ¾*** **Qualify Description**

Data Type: CHAR

Format: A(50)

Description: The qualify description field is used to define each qualifier code. Please refer to the “Qualify” table for codes and descriptions.

Notes:

Validation:

Chapter 4

The Outlier Table

The outlier table is used to identify records with highly unlikely use rates (outliers). Each row in this table corresponds to either one pesticide application for production agricultural reports or a monthly summary for other uses. Reports of applications for any use other than production agriculture only include the total of all uses in a month for each pesticide, site treated, and applicator. The type of report is identified in the UDC table by the field RECORD_ID. Production agricultural reports have RECORD_ID values of 1, 4, A, or B; monthly summary reports have record_id values of 2 or C. Each row is uniquely identified by the column use_no that occurs in both the UDC table and the outlier table. The other three columns in the outlier table contain the flags for the three different criteria. A value of Y in one of these columns indicates that the rate is an outlier by that criterion. A value of N indicates it is not an outlier by the criterion. A blank or space indicates that the criterion could not be applied to that particular record. If no criterion applies to a row in the UDC, there is no corresponding row in the outlier table.

The first criterion column in the outlier table, AI_A_1000_200, flags records with rates higher than 200 pounds of active ingredient per acre (or greater than 1000 pounds per acre for fumigants). The second column, PRD_U_50M, flags rates 50 times larger than the median rate for all uses with the same pesticide product, crop treated, unit treated, and record type (that is, production agriculture or monthly report). The third column, nn4, flags rates higher than a value determined by a neural network procedure that approximates what a group of 12 scientists believed were obvious outliers. These criteria are explained in more detail in the data dictionary below.

Although applications or rows are flagged, the only values tested are rates. Thus, there is no reason to believe that the other data in a row, such as time and location of the application, are not correct. Also, note that rate is not one of the fields in the UDC table. Rates are calculated by dividing the pounds of pesticide used by the acres or unit treated. Thus, an extremely high rate value could occur from either an extremely high pounds used or extremely low unit treated.

Only extremely large rates are flagged, not extremely small ones, because only large values will have a major influence on statistics involving pounds of pesticide use. What value to use for the maximum rate in each criterion is somewhat arbitrary; the value determines how conservative one wants to be. We chose maximum rates to be close to what were considered obvious outliers by a group of scientists in a survey described below in the description of the neural network criteria.

There are many possible methods for determining if a value is an outlier. If we knew the maximum label rates for particular uses, then rates in the PUR could be compared to these maximum rates, but unfortunately this information is not available in the PUR or in the Pesticide Label Database. The other methods to identify outliers involve looking at the distribution of the actual use rates. If the values are normally distributed, then one can identify outliers using a number of statistical procedures. If the values have an unknown or nonstandard distribution, then there exist no standard statistical procedures for identifying outliers. Nevertheless, people can look at a distribution and usually say with different degrees of confidence whether some value is an outlier. This suggests there should be some kind of procedure that can be developed to make similar judgments.

For most of the pesticide use data, distributions of rates are not even close to normal. They may have several different peaks (multi-modal). They can have either very broad distributions or very narrow distributions. None of the standard statistical measures of outliers are very useful for these data. The best single method is the one based on neural networks. However, each different criterion will catch different outlier values so it is usually best to use all three criteria. It should be noted that these criteria are not perfect. They are conservative, meaning a value must very extreme to be flagged and so they will miss some errors. On the other hand, they may occasionally flag an extreme value that is actually correct. Because the criteria are conservative these later kinds of errors are minimized.

For a more detailed explanation of the procedures used to identify outliers, see the report "*A Computer Program to Identify Outliers in the Pesticide Use Report Database*", L. Wilhoit, April 1998, DPR report PM 98-01.

Outlier Table Record Structure and Data Dictionary

Field Seq. No.	Field Name	Type	Mask
1	USE_NO	N	99999
2	AI_A_1000_200	C	A
3	PRD_U_50M	C	A
4	NN4	C	A

Outlier Table

Field Sequence No. ***1***

Field Name: ***USE_NO ¾*** **Use Number**

Data Type: NUMERIC

Format: 9999999

Description: System assigned sequential number to uniquely identify a pesticide product use record within a year.

Notes: Corresponds with USE_NO in the UDC table. If no criterion applies to a row in the UDC table, there is no corresponding row in the outlier table.

Outlier Table

Field Sequence No. ***2***

Field Name: **ai_a_1000_200 ¾ Criterion 1 ¾ Pounds per acre of active ingredient is larger than 200 (for non-fumigants), or 1000 (for fumigants).**

Data Type: CHAR

Format: A

Description: Records were flagged by criterion 1 if the pounds per acre of a non-fumigant active ingredient were greater than 200 or if the pounds per acre of a fumigant active ingredient were greater than 1000. These limit values were chosen based on what is known about typical rates of use for most pesticides. Note that this criterion uses the pounds of active ingredient. Also, this criterion only applies to records where the unit treated is acres.

The other two outlier criteria use pounds of pesticide product and apply to any unit treated, such as square feet or cubic feet.

Notes: Y value indicates that the rate is an outlier by this criterion. N value indicates it is not an outlier by this criterion. A blank or space indicates that the criterion could not be applied to that particular record.

Outlier Table

Field Sequence No. ***3***

Field Name: **prd_u_50m ¾ Criterion 2 ¾ Pounds per unit treated of a product is larger than 50 times the median.**

Data Type: CHAR

Format: A

Description: Records were flagged by criterion 2 if the pounds of pesticide product per unit treated were greater than 50 times the median value of all rates with similar types of use. The median, like the mean (average), is a measure of the location of a set of values and is defined as the value in the set that has an equal number of values above and below it. It was used rather than the mean because it is not as likely to be affected be a few extreme outliers. The median was calculated from the set of all use rates of the same pesticide product and uses as that of each record being examined. By the same uses, we mean the uses of a product on the same crop or site, same unit treated, and same record type. A record type is either a production agriculture report (which includes a single application) or a monthly summary report

Notes: Y value indicates that the rate is an outlier by this criterion. N value indicates it is not an outlier by this criterion. A blank or space indicates that the criterion could not be applied to that particular record.

Outlier Table

Field Sequence No. 4

Field Name: **nn4 ¾ Criterion 3 ¾ Pounds per unit of product is larger than a value generated using a neural network.**

Data Type: CHAR

Format: A

Description: Records were flagged by criterion 3 if the pounds of a pesticide product per unit treated were greater than a limit value that was calculated using a neural network procedure.

A neural network is a function that maps a set of input values to a set of output values. This function has a large number of parameters that must be determined so that the function will give the correct outputs for every possible set of inputs. The values for these parameters are found by a training procedure that involves presenting to the neural network program data consisting of many sets of input and corresponding output values. The program then adjusts the parameters in the neural network function until it produces the correct output values for each input set. Once the neural network has been successfully trained, it can then be used to produce appropriate output values for any input data set provided to it.

The data used to train the neural network used in the PUR outlier program were generated from frequency distributions of the pounds of pesticide product per unit treated for a selected set of pesticides and sites. Groups of pesticides and sites were chosen that included a wide range of types of distributions, including many unusual distributions. Two hundred frequency distributions were plotted and then these plots were examined independently by 12 scientists in DPR who marked rates on each plot they thought were outliers.

The results of this survey were summarized by finding an outlier maximum rate for each distribution. The maximum rate was set at a value where all 12 scientists thought higher rates were obvious outliers. These maximum rates were used as the output values for training the neural network. The input values were a set of statistical measures that described the frequency distributions.

These sets of input and output values were used to train the neural network. After the neural network was successfully trained, it was used to find the outlier maximum rate for all sets of pesticide use types in the PUR.

Notes:

Y value indicates that the rate is an outlier by this criterion.
N value indicates it is not an outlier by this criterion. A blank or space indicates that the criterion could not be applied to that particular record.