WastePro Memorandum

To: Richard Gladstein (ENRD), Department of Justice

From: Jeff Karnes

Date: December 18, 2006

Re: Hazardous Waste Receipts into MSP 1992 through 1996: Report Development Process

---

Introduction:

The purpose of this project is to produce a list of the generators that sent hazardous waste into the Marine Shale Processors Inc (MSP) plant in Amelia, LA, USEPA ID: LAD98105776. The data was summarized by year for the 500 largest generators for 1992 through MSP's shutdown in 1996. Originally, 1991 was to be included, but complete data was not available in an electronic format. This memo details the process used to convert the original data into a relational database.

Data Source:

Hazardous waste shipments are tracked by a manifest system. The manifests contain data sections that provide information regarding the identity of the generator of a waste, shippers that take possession of the waste while in route, and the receiver of the waste. For each individual waste stream in the shipment, the manifest contains a set of information that includes a descriptive name for the waste, a shipping name per DOT Hazardous Materials Regulation 181, and quantity by container or standard measure. For every waste, each applicable EPA hazardous waste code must be recorded. Additional fields identify the form of the waste (e.g., liquid, solid) and the expected next processing method for the waste.

The manifest travels with the waste shipment and acts as a chain of custody document. Each signer of the document maintains a copy. When the manifest is received by the destination facility, a copy is sent to the state environmental office. The state then forwards a copy of the manifest with all signatures to the original generator.

The Louisiana Department of Environmental Quality, LADEQ maintained copies of waste manifests received by MSP. Sometime in 1991, LADEQ began entering the manifest information into an electronic database. Not all states follow this practice. Louisiana, California, Texas,

New Jersey, and New York are among the most active states in managing manifest information electronically.

To determine the waste receipts into MSP from 1991 through its closure in 1996, a request was made to the LADEQ Office of Public Records for an extract of this information, Public Records Request #0022398. In response to YK2000 issues, LADEQ changed manifest systems, but did not convert historic data to the new system. The raw data had to be obtained from archived data tapes. James D. Miller, Environmental Scientist Manager, Compliance Assistance Section, LADEQ created a custom extract from the tapes.

For each manifest, the following information was obtained: generator EPA ID, the EPA ID of each transporter, facility EPA ID, waste quantity, unit of measure, handling code, form code, and hazardous waste codes with waste code descriptions. The data was provided on CD as two Microsoft Excel workbooks containing a combined 28 worksheets. Each worksheet contains a maximum of 65,000 lines of data

The archival tapes did not contain the site identification data; such as, facility names, addresses, and contacts. This information changes over time as operations are sold and personnel change. If the same basic business activity occurs continuously at a site, its EPA ID remains constant. The EPA maintains a current database with the most recent information associated with each ID. Over the course of this project, a list of generator and transporter EPA IDs was created. This information was provided to EPA Region VI. Buddy Jones, RCRA Info Regional System Administrator matched the IDs to the EPA database and provided the current facility name and address information. This information was delivered in two Excel worksheets via email.

Development of Relational Database:

The data was delivered in a flat file format. In this format, each unique combination data for a manifest waste stream is recorded as an individual data record. For example, three records would represent a waste stream with three hazardous waste codes. If that shipment had been handled by two transporters, then the above stream would expand to six records. This format leads to extremely large files with vast amounts of redundant information. Some individual manifest could contain up to 50 records. It is difficult to create useful accurate information from data in this format.

The data redundancy problem is addressed by converting the data into a relational database. A relational database organizes the data into a system of tables that are linked according to key variables. The result is a streamlined system with minimal redundant information. A major portion of this project involved converting the data into a set of relational data tables.

Step One: Import Data

The first task was to import the individual Excel worksheets into data tables in Microsoft Access. Access is a powerful relational database system suitable for projects of this size. All the individual tables for a year were appended to form a complete annual table for each year.

Step Two: Create a Uniform Volume Field

The hazardous waste manifest allows volume information to be presented using a variety of units. To evaluate the information, these must be standardized to a single unit of measure. Tons were chosen as the standard for this project. A new equivalent tonnage was created and populated according to a series of rules. The pounds to tons conversion is defined, but others require an estimated conversion. The majority of material with volume provided in gallons was assumed to be bulk liquids based on the observation that many manifests have volumes around 5,000 gallons. This is the most common size of an over the road tanker. MSP had a steady need for fuel grade liquids to maintain the kiln temperature. These wastes would have been mixtures of light hydrocarbons. These have a density of around 7 lb/gal. MSP did not burn much wastewater or chlorinated streams. These have densities of 8.3 and 10 lb/gal, respectively. Consequently, a conversion factor around 7 lb/gal times 1 ton/2,000 lb was used. When the volume of material was recorded in cubic yards, it was assumed to be a debris stream. This is validated by the fact that much of this material is shipped in dump trailers, roll-offs and boxes. Soil shipments are typically recorded as tons. The density of this material varies greatly. For a first pass, a density of 0.8 tons per yard was assumed.

After the initial conversion, the data was sorted by the converted tons. Any load record above 30 tons was inspected. Generally, a truck shipment of hazardous waste contains a maximum of 20 to 22 tons. This is based on a DOT weight standard for the combined trailer and load for the typical van (the ubiquitous 18-wheeler). Very large records were examined to evaluate potential data entry problems. A common problem occurs when pounds are recorded as gallons. For example, a manifest lists 40,000 gallons of material. Records of this type were assumed to be pounds and corrected. In other instances, a record might contain an obvious wrong value due a typographical error. These were investigated by looking at the container type and other manifests from the same shipper. This inspection process did correct many of these. Some streams initially recorded in yards were too large after conversion. These were assumed to be less dense debris and the conversion factor was lowered to 0.5 ton/yd.

The exception to the 30-ton limit is railcars of bulk liquids. These are typically 20,000 to 30,000 gallon containers. These would convert to 70 to 105 tons. Rail shipments tend to be one of many from a handful of generators. Additionally, the universe of railroads is small and they often require multiple rail lines. Apparent rail shipments were not modified.

Utilizing the query features of Access, totals for each year were compiled to check for reasonableness. Data for 1991, 1993, and 1995 was compared against data published online for EPA biennial reports (BRS). Data for 1991 was found to be incomplete and a number of months appeared to be missing. For 1993 and 1995, the data was found to be 3 to 8 thousand tons or about 4 to 9 percent less than the BRS data. This was not unexpected. The BRS data is derived

from the same manifest information and often contains the types of errors that were corrected for in the new database. This conclusion is based on extensive work with manifest and BRS databases. The even years were evaluated for consistency with the odd years.

Table One: Annual Database Totals versus BRS Totals

Year	Database Total (tons)	BRS Total (tons)
1991	9,430	80,106
1992	100,.499	
1993	83,498	91,658
1994	74,215	
1995	72,745	75,662
1996	29,174	

Step 3: Identify Relational Tables and Links

Next began the process of identifying data that could be isolated in separate tables to minimize redundancy. This requires that for one field in a table each record is unique. This field often has to be a construct of two or more fields. On a hazardous waste manifest, multiple waste streams can be listed. Each may differ by container, waste code, form, or another variable. Analysis determined that less than one percent of the manifests contained multiple streams. Further examination of this subset revealed that the variations were by container and form, but the waste codes were all the same. This analysis is not concerned with containers and form. Consequently, the multiple streams were consolidated into a single manifest stream. The manifest number became the key variable for most of the data tables.

A table containing unique manifests along with equivalent tons and several other variables was created for the 1992 data. Next, tables of unique manifests and generators and of unique manifests and transporters were created. Additionally, a table of manifests to waste codes was prepared. This remained a large table because each manifest/waste code combination required a record. The waste descriptions in the data were simply standard description for each waste code. A database of unique waste codes and descriptions was created.

This process involved utilizing the statistical query features of Access and utilizing interim tables. In order to ensure that no data was lost and no redundant data was retained, the results of the process were tested against a set of manifests with multiple containers, shippers, and or waste codes. The results were proven to be valid.

This process was repeated for each year of data including 1991.

Step 4 Consolidate Data Tables

Next, the results of Step 3 had to be exported as tables into a clean database. This had to be done the relational nature of Access maintains linkages with the original redundant data table. This links confounds any further use of the data. In the new database, the individual annual tables were consolidated into single tables for all years. The generators, transporter, and waste description tables were modified to eliminate any redundancy.

Step 5: Establish Relationships

At this time, the links between tables were defined. A record in the volume table links to a single record in the generator table. A record in the volume table may link to multiple records in the waste code table by manifest. A series of queries were created to form the foundation for the desired reports. Access requires a sequential progression to assemble the data.

Step 6: Resolve Multiple Waste Code Problems

Each waste code associated with a stream created a record in the database. Waste streams will generally require 1-5 EPA waste codes. A common combination of codes is for a mixture of chlorinated and non-chlorinated waste solvents exhibiting characteristics of flammability along with one or two heavy metals. This results in five applicable codes. Since the vast majority of waste into MSP came from the hazardous waste management industry, two types of streams could readily have twenty or more waste codes. Lab packs contain small quantities of unused chemicals and waste laboratory quantities in an over-pack. Individual lab packs often require more than 20 codes. These facilities consolidate material from numerous generators in a single bulk container. All the codes associated with those streams become attached to any waste removed from the container. Furthermore, once a code enters a container, it is required to be added to any waste taken from the container until the container is emptied to RCRA standards. This is referred to as the "derived from" rule. This leads to large numbers of waste codes on the manifests that do not really describe the waste.

The solution to the problem is to define a field for each waste code. Through an iterative process, the first waste code for each manifest was captured in a first waste code field and moved to a new data table. This process was repeated until the new table contained the manifest number and the first eight waste codes. Useful data lost from lab packs is minimal because the volumes are very small. Data lost from others is negligible because the codes have little meaning for many TSDF tank streams. Since the waste codes were alphabetized, almost all characteristic, D-codes were retained. The original waste code table was retained. Data table relationships were updated.

Step 7: Attach Names and Addresses

As discussed in the data source section, the LADEQ data contained US EPA IDs, but lacked the generators names and addresses. The data obtained from EPA Region VI was converted into Access data tables and linked to the system

After merging the LADEQ derived data with the EPA identification information, 35% of the 6,850 generators were still missing names. This large number of generators was associated with only 2% of the waste volume. Based on years of experience working with manifest databases, the data was inspected for common problems; such as, transcribed digits. This process resulted in over 200 records being assigned to generators of known identity in the database. Most of the remaining sites without names have non-standard EPA IDs.

Most EPA IDs begin with a two-letter state abbreviation followed by a D. For transporters, the third letter is sometimes T. For are large number of records, the third letter is P or Tmp replaces the third letter and digits. These designations are used for small quantity generators or one-time generators. Generally, there is only one manifest associated with each of the numbers. In a few instances, one of these numbers is aligned with multiple shipments. This is probably due to a large spill or cleanup conducted by a government agency. About 16 cases like this occur in the data. GATMP0002010 at 1,671 tons accounts for over one-half the waste in this category. If necessary, each of these can be researched at the state level.

The process was repeated for the transporters. Initially, a similar percentage was missing names. Since the transporter universe is smaller, the cleanup process was even more successful for this group.

Step 8: Create Reports

Utilizing the Access report features, three types of reports were created for the data covering 1992 through 1996. To facilitate this a ranking field the ranks generators by total waste shipped for the period was created. The entire universe of generators numbered 6,636 and shipped 360,000 tons. Reports for the entire universe would be cumbersome and most of these generators had shipped well around a truckload, 20 tons. Limiting reporting to the Top 500 was determined to yield a more manageable report with a minimal lost of data. This cutoff captures 96.5% of the total volume and the 500th generator is at about 20 tons.

The following reports were created:

MSP Waste Receipts - Top 500 Ranked: ranked by total waste from 1992 to 1996
This report has been delivered to DOJ under separate cover.
MSP Waste Receipts - Top 500 Alphanumeric: listed by USEPA ID
MSP Waste Receipts - Top 500 Annual: listed by USEPA ID with subtotals for individual years.

The reports can be viewed in Access. This requires that the user have a valid copy of Access. The data tables are not protected in this situation and can be corrupted. At this time, it was decided to move the report output to Adobe Acrobat .pdf format. These files can be viewed with the Acrobat reader available free on the internet. They cannot be modified. However, they be printed and are searchable.

Future:

The data system contains all the information at the individual manifest with waste code levels. Future reports and/or queries can be developed to view this information. This could be useful in future discussions with generators. It is not known if there any waste code specific issues relevant to this period.