Journal of Hazardous Materials, 17 (1988) 287-304
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

TOWARDS IMPROVED HAZARDOUS MATERIALS FLOW DATA*

GEORGE LIST and MARK ABKOWITZ

Department of Civil Engineering Rensselaer Polytechnic Institute Troy, NY 12180 (U.S.A.)

(Received August 31, 1986; accepted in revised form June 3, 1987)

Summary

With the increasing traffic volumes of hazardous materials, risk minimization has become a topic of widespread interest to the transportation community. Government and industry alike see a need to plan carefully the movement of these dangerous substances over the world's network of highways, railroads, waterways, and other transportation facilities.

Flow data can be of great value in planning the minimum risk movement of these goods. For example, local governments can use the data to help set priorities for emergency preparedness training. Since response teams typically cannot afford to be prepared for every conceivable emergency, flow data can tell them which types of substances they are most likely to encounter. This in turn can help direct training programs and equipment purchases. At higher levels of government, decisionmakers can use the data to weigh the benefits and costs of route control, inspection programs and other actions.

This paper discusses the challenges involved in developing such a database for the United States (U.S.). It focuses on issues of completeness, consistency, and compatibility, both between and within the source databases commonly available. Suggestions, then, are given for altering these databases so they can be more useful for such purposes.

Introduction

With increasing traffic volumes of hazardous materials, risk minimization has become a topic of widespread interest to the transportation community. For example, within the past few years, the U.S. has seen an international conference [1], a detailed federal investigation [2,3], and numerous risk management studies [4–6]. Government and industry alike see a need for safety and policy analysis to plan the minimum risk movement of these dangerous substances over the world's network of highways, railroads, waterways, and other transportation facilities.

Flow data can be of great value in this planning process. For example, local governments can use the data to help set priorities for emergency preparedness training. Since response teams typically cannot afford to be prepared for every conceivable emergency, flow data can tell them which types of substances they are most likely to encounter. This in turn can help direct training programs

0304-3894/88/$03.50

and equipment purchases. At higher levels of government, decisionmakers can use the data to weigh the benefits and costs of route control, roadside inspections, site-specific data gathering efforts, representative analyses of risks by mode, and other actions.

This paper discusses the challenges involved in developing such a database for the United States (U.S.). It focuses on issues of completeness, consistency, and compatibility, both between and within the source databases commonly available. Suggestions then are given for altering these databases so they can be more useful for such purposes. While the paper is based on experience in the U.S., researchers in other countries may find the observations useful in formulating improvements to their own databases.

The data needs

Conceptually, a hazardous material flows database needs to include each shipment's origin, destination, date of shipment, mode, commodity, weight, and distance. Moreover, because of the way in which the risks are determined, there is a need to know the type of vehicle and container employed and the route used.

To illustrate the structure of such a database for the U.S., the following definitions could be used for the data items. Postal delivery zones (i.e. ZIP codes) could be used to record the origin and destination (another possibility would be Standard Point Location Codes [17]). The date could be given as month and year, although day within the month would also be helpful. For mode, shipments involving several modes could be broken down into separate records, one for each leg of the trip. A flag could be used to indicate whether only one mode was involved or several. The shipment legs need not be cross-referenced. For the commodity field, Standard Transportation Commodity Codes [8] could be used, or UN/NA codes, or any of a number of other possibilities. Weight should be given in consistent units, such as pounds, tons or kilograms. Generic vehicle and container types should be used, along with a cross-reference table for specific makes and models.

To record the route that was followed, major landmarks could be used. For rail shipments, these landmarks could be the junction points presently used for the Waybill Sample [9]. For multiple-line shipments, no changes would be necessary; for single line (single railroad) shipments, where the data are presently missing, equivalent internal "junction points" could be used. For truck, sequences of highway "numbers" could be used, each identified by letters (e.g. I for interstate, US for U.S. highway, and state codes such as PA, IN, and IL for state highways and a number; the result would be I78, US22, PA320). These fields could be left blank for shipments that only use local roads; and for long distance shipments, the number of highways listed could be limited to six or seven, focusing on the first, the last, and a logical group of intermediate ones.

Potential sources of data

For developing a hazardous material flows database such as the one described above, several source databases are presently available in the U.S. The major ones [3,10] are listed in Table 1. They range in scope from the broad-based Commodity Transportation Survey (CTS), which contains data on shipments of all domestically manufactured items from point of manufacture to first destination, to the narrowly focused "EPA" database, which is only for hazardous waste shipments, primarily by truck.

Considerations of space here do not allow a complete review of all the databases. The paper will focus only on the three most widely available: the Waybill Sample (1972–1983), the Waterborne Commerce Statistics Center database (WCSC, 1977 and 1982), and the Commodity Transportation Survey (CTS, 1977).

Defining the hazardous commodities

Any researcher who wishes to develop a hazardous materials flow database must first decide which commodities to include. In the U.S., the Secretary of Transportation has developed a list of over 2,400 materials in various quantities and forms (Hazardous Materials Transportation Act) and by regulation (49 CFR 171.8) that may pose an unreasonable risk to health and safety, or to property, when transported in commerce [11]. Inclusive in the list are several substances and wastes classified as "hazardous" in order to coordinate DOT's regulatory program with that of the Environmental Protection Agency (EPA). The primary reason for designating these materials is their long-term effects on health and the environment [12]. For each substance, EPA has established a "reportable quantity" (RQ) which indicates the quantity and concentration of a chemical that could pose a threat of pollution. RQs for most substances are one pound, although EPA is currently studying the effects of changing the RQ level [13]. Packages containing more than the RQ of the hazardous substance are subject to DOT regulation. DOT regulations also apply to hazardous wastes that are subject to EPA's manifest system under the Resource Conservation and Recovery Act.

It is a challenge, however, to use this list of commodities in its present form. Its numerical codification, which is maintained by DOT's Office of Hazardous Materials Transportation (OHMT), is not cross-referenced to any other commodity classification scheme.

Fortunately, another option is available. The Association of American Railroads, at the direction of the Interstate Commerce Commission, maintains a set of "49-series" codes. Part of the Standard Transportation Commodity Codes (STCC), the "49-series" provides a seven-digit numerical identification for each commodity (two of the seven digits are "49"), a hazard class designation

TABLE I

Commodity flow databases

Database	Kept by	Years	Modes	Commodity codes	Conversion table	Strengths	Weakness/ Drawbacks
Commodity Transportation Survey	Bureau of Census	1977	all	5-digit STCC	yes	• multi-modal • consistent selection procedure for all sample data points for all modes • cross-checked against the Census of Manufacturers	• only 5-digit level of detail on commodities • no Haz Mat flags • only shipments from manufacturing sites to first destinations • only "principal" mode is reported
Truck Inventory and Use Survey	Bureau of Census	1977, 1982	highway	simple classes	no	• cover all trucks used in the U.S. • contains Haz Mat-related data items • sample biased toward heavy trucks	• no flow data • only rudimentary commodity information • tractor database, not a trailer database-reflects tractor use not trailer use
Motor Carrier Census	Bureau of Motor Carrier Safety, FHWA	most recent five years	highway	hazard classes	no	• comprehensive listing of carriers and truck fleet operators	• no flow data • mileage and fleet size data are sparse
TRANSEARCH, FREIGHTSCAN, etc.	Consulting Firms	varies	all	varies, up to seven-digit for rail	yes	• cross checked against other production/consumption data • melding of the best available for each mode	• truck flows predominantly based on CTS data (see above) • not in the public domain

National Motor Truck Database	Consulting Firm	1977 to present	highway	varies, up to seven-digit STCC	yes, where commodity code is provided	• focuses on long-distance highway flows • true flow data • describes the vehicle used to carry the commodity	• purposely excludes short-haul truck movements, especially in the Northeast • not in the public domain
Waybill Sample	Interstate Commerce Commission	At least past 12 years	rail, TOFC/COFC	Seven-digit STCC	yes	• well-organized sample (1%) of all rail flows • database is consistent enough to allow trend analyses • contains some routing information	• The list of Haz Mat STCC Codes varies from year to year
Waterborne Commodity Statistics	Army Corps of Engineers	At least 12 years	water, domestic and international	four-digit WCSC code	only to a limited extent	• "100%" sample of all vessel movements • complete routing information	• only 163 commodity codes in all, so level of detail is weak • conversion table has some incorrect cross-references
TRAIN II	Association of American Rail-roads (Railinc)	current	rail, TOFC/COFC	seven-digit STCC	yes	• 100% data on all movements for participating railroads • routing information	• not specifically designed to record car movement histories • not in the public domain
Hazardous Waste Shipment Data	States, for the EPA	varies	primarily highway	either EPA codes or OHMT	no	• "100%" sample of all hazardous waste shipments • actual flow data	• many states do not computerize the data • no consistency to commodity code usage • no routing information

(hazard classes are the primary commodity groupings employed by OHMT; they are based on the commodity's most dangerous attribute), a UN/NA number, and a textural definition. A corresponding numerical translation table allows conversion of these 49-Series STCC's into regular STCC's; for example, 28-series for chemicals and 29-series for petroleum products.

Working with the waybill sample

As may be the case in other countries, the U.S. maintains a database specifically for rail shipments. It is a stratified sampling of the shipments made by U.S. railroads. The Waybill Sample is collected every year, per direction of the Interstate Commerce Commission, with the rate of sampling ranging from 1% for single car shipments up to as much as 50% for large multiple-car shipments. (A waybill is a contract for the movement of freight.)

For 1984, the sample contains records on approximately 335,000 shipments. The origin (city and state), destination (same basis), commodity (seven digit STCC, including 49-Series codes, where used), number of cars, and shipment weight are all given as well as the railroad junctions traversed for multiple railroad shipments.

The sample comes very close to meeting all the needs of a hazardous materials database. It lacks only route data for the shipments involving one railroad. Moreover, the data appear to be representative of the actual traffic flows. This is based on a comparison of the 1983 and 1984 data with actual shipment records from three major chemical companies [2].

The major challenge in using the Waybill Sample relates to commodity selection. In principle, one only needs to work with the 49-series commodity codes. But, if a multiple year database is desired, the situation is not that simple. The list of 49-series STCC codes differs between years, and prior to 1976 it did not exist. If only the 49-series codes are used, the number of hazardous material shipments in the Waybill Sample appears to rise from none in 1975 to over 11,000 in 1983, as shown in Table 2. Hence, one needs to generate an all-inclusive list of hazardous material commodity codes based on the 49-Series codes as well as the numerically cross referenced standard 28-Series (chemical) and 29-Series (petroleum) codes.

One potential difficulty with this all-inclusive list is that there may be substantial non-hazardous traffic volumes moving under these regular codes, particularly in recent years. (In principle, all hazardous material traffic moves under 49-series codes.) This is not a trivial problem, because as Table 2 shows, if the 28- and 29-series flows are included in the universe of harzardous materials flows, while the percentage of 49-Series records has continued to climb, in 1983 only three of every four of these "hazardous material" shipments used a 49-series code.

Another challenge when including the regular 28- and 29-series codes is the

TABLE 2

Trends in hazardous materials statistics based on the waybill database

Year	Hazardous total records	Materials records			Flow trends		
		49-Series	Regular	Total	Cars (000)	Tons (000)	Ton-Mi (mill.)
1972	207,729	0	11,388	11,388	1,190	67,066	34,478
1973	213,109	0	11,800	11,800	1,234	71,159	40,678
1974	198,784	0	11,204	11,204	1,172	70,476	41,318
1975	173,528	3	10,056	10,059	1,041	64,449	38,427
1976	182,264	6,394	4,829	11,223	1,172	72,815	44,736
1977	180,781	6,561	4,516	11,077	1,147	73,297	44,604
1978	185,350	6,773	4,015	10,878	1,124	73,165	45,212
1979	194,234	7,157	4,001	11,158	1,144	76,650	47,711
1980	181,558	6,811	3,570	10,381	1,068	73,504	45,329
1981	208,580	0	12,440	12,440	977	69,726	43,394
1982[a]	162,210	7,974	3,550	11,514	852	58,095	—
1983	236,615	11,093	4,594	15,687	1,077	73,161	53,269

[a]The data for this year appear to be unreliable

identification of hazard classes for each of these commodities. The 2,466 49-Series STCC's translate into only 986 regular seven-digit STCCs, reflecting the greater level of specificity provided by the 49-series codes. Almost a quarter of them (544) translate to just one regular STCC, namely "chemicals, not elsewhere classified" and another quarter translate into two other regular STCCs. Only 637 have a unique correspondence. Because of this, while the 49-Series codes do have unique hazard classes, the "hazardous" regular STCCs typically have two or more. In fact, one has 14, another has 7, two have 6, eight have 5, two have 4, twenty have 3, and 194 have two. The fortunate part is that the traffic moving under these codes is less than 1% of the total [2].

Working with the WCSC database

The U.S. Army Corps of Engineers collects data on all vessel movements on U.S. waters [14]. The resulting Waterborne Commerce Statistics Center (WCSC) database contains complete data on all waterborne hazardous material flows in the U.S. But it is a significant challenge to work with the database. Aside from the fact that it is a "100%" sample and hence voluminous; the commodity definitions are very broad, with only 163 codes in all, 49 of which pertain to hazardous materials.

Based on experience in selecting the hazardous materials codes, researchers should take note that such lists for this database are best developed by hand. In the OTA study [2], an attempt was made to develop the list using cross-reference tables, but it met with only limited success. The starting point was the 49-series STCC's described earlier. Through a process involving several conversion tables, these were translated into 986 seven-digit STCCs, 89 four-digit STCCs, 115 four-digit Standard Industrial Classification (SICs), 2,978 Schedule A codes (used by the Bureau of the Census for imports), and 1,572 Schedule E codes (used by the Bureau of the Census for exports). The final Schedule A and E codes yielded 100 WCSC codes.

Unfortunately, there is little rationale to these 100 codes. While all petroleum and coal products and almost all chemicals are identified, so are raw cotton, live animals, clay, vegetables, animal feeds, furniture and fixtures, electrical machinery, and a host of other equally "non-hazardous" commodities. The difficulty is that the conversion tables do not provide enough specificity.

Moreover, some commodities are missing; such as sodium hydroxide and sulfuric acid. They are not cross-referenced correctly with their corresponding WCSC codes. In the case of sulfuric acid, the Schedule A and Schedule E codes are both cross-referenced WCSC codes other than sulfuric acid. In fact, the WCSC code for sulfuric acid does not appear in either the Schedule A to WCSC or Schedule E to WCSC conversion tables.

Fortunately, the flows developed from the manually selected codes are a reasonable reflection of what one would think the flows should be. They follow the distribution of petroleum; perhaps not surprisingly, because 85% of the tonnage is either crude or processed petroleum.

Working with the Commodity Transportation Survey

The third major U.S. source of commodity flow data, and the only one for truck and air, is the Commodity Transportation Survey (CTS). The most recent CTS, for 1977 [15], contains data on shipments of manufactured goods from point of manufacture to first destination. Each record provides the tonnage, origin (state or production area), destination (same basis), commodity code (up to five-digit level Transportation Commodity Code (ICC) – the same as a five-digit STCC), principal mode of transport, weight block, and value block. The data is based on voluntary responses from approximately 16,000 of the 19,500 establishments to which survey forms were sent. It is checked against the Census of Manufacturers Survey using value of shipment as the standard of comparison to ensure that the expanded value of shipments corresponds "closely" to the value of the commodities produced.

The version of the CTS available on magnetic tape is actually many databases in one. It contains a database for each combination of (1) geographic clustering (i.e., state or production area), (2) degree of commodity specificity (none, two-digit, three, four, and five-digit TCC, and (3) type of flow data (i.e., from one location to another or just to a given location, e.g., flows to the state of NY regardless of origin).

TABLE 3

Hazardous material ton and ton-mile statistics: 1977 State-to-State five-digit CTS flows versus 1977 Destination State five-digit CTS flows

Mode	State-to-state flow data		Destination state flow data	
	Tons (000)	Ton-miles (millions)	Tons (000)	Ton-miles (millions)
Truck	133,576	28,399	283,331	54,386
Rail	82,428	49,763	147,025	77,684
Water	202,754	105,175	260,898	120,733
Air	53	63	129	163
Total	418,811	183,400	691,383	252,966

When developing a hazardous material flows database from one of these CTS databases, there are some important facts to consider:

- The number of responses in each database is inversely related to the level of detail. Thus the completeness of the data diminishes as the level of detail increases. The reasons for this seem to be: (1) the Bureau is under obligation to protect its sources and (2) respondents are not consistent in the commodity detail they provide.
- The databases are not cross-checked to ensure that totals from one database closely match those from any other.
- The production area-based databases reflect only flows to and from production areas; all other flows, namely those that originate or terminate outside these areas, are omitted.

To illustrate the implications of these facts, consider the differences between the five-digit level destination state (DES) database and the five-digit level state-to-state (STS) database. By definition, the first has records for flows to a given state, while the second has records for flows from one state to another. Though one would think that the total flows to a state from the destination state database should match the to-state totals from the STS, Table 3 shows that there are consistently greater totals for the DES database than for the STS database, often by a factor of two to three. Furthermore, Table 4 shows that one can develop two different estimates of the tons being shipped by commodity and mode.

Another fact to consider is that the commodity specificity is only at the two, three, four or five-digit STCC level. The CTS does not have a seven-digit level database, nor does it have a field for a hazardous materials flag. Assumptions, then, must be made when selecting commodities for a hazardous material flows database. In work done for the OTA study, a list of codes was developed using the 49-series codes described earlier. The 49-series codes were converted to

TABLE 4

Principal commodities for truck based on the 1977 five-digit CTS database

Commodity	Tons (000)	
	STS	DES
Chemicals (total)	59,695	148,714
28194	8,800	10,068
28211	6,731	10,540
28182	2,957	3,509
Petroleum products (total)	64,793	111,270
29111	23,166	40,594
29116	11,173	19,593
29113	9,024	19,691
All other (total)	9,088	23,348
20841	1,714	2,010
20851	1,014	2,009
20144	1,013	2,293
Total	133,576	283,331

Note: Top three based on STS tons; STS: State-to-state flow database; DES: Destination state flow database.

regular STCCs (using the numerical cross-reference tables) and then shortened to five digits.

For the resulting 147 five-digit codes, the five-digit CTS state-to-state database has flows for 98. Of these, 47 are chemicals, twelve are petroleum products, and 39 are other commodities. Of the 39 "other" commodities, the 16 shown in Table 5 are probably not hazardous, because they are items such as "commercial refrigeration equipment" and "motor vehicle accessories". While they have a logical basis for being categorized as hazardous commodities, researchers should probably delete the data for them when creating a database since very little of the actual tonnage moved is hazardous.

Several other points are equally important. The CTS' focus on "first leg" shipments misses significant portions of the overall truck-based flows. For example, local truck deliveries of petroleum products are missing. The STS database does not show any trucks at all carrying petroleum products in the South Atlantic, West North Central, or New England regions.

This means that for truck-based flows, the CTS should be supplemented with other sources. For example, for the OTA study, data was collected by the American Petroleum Institute (API) from its member oil companies. API asked each respondent to provide a profile of the distribution patterns at its terminals, categorized by type of product and type of delivery (e.g., proprietary fleet

TABLE 5

Highlights of the CTS five-digit hazardous commodity list

Basis	
2466 seven-digit 49-series STCCs	
Initial list	
986 seven-digit regular STCCs	
Intermediate list	
147 five-digit regular STCCs	
Deletions from the 98 five-digit commodities actually in the CTS	
20421	Prepared feed..other than pet food
20713	Chocolate or cocoa products or by-products
21111	Cigarettes
24999	Wood products, nec.
33129	Primary iron or steel products, nec
34439	Frabricated plate products, nec
35339	Gas or oil field machinery or tools, nec
35511	Dairy or milk product plant mach. or equip.
35611	Industrial pumps, pumping equip, or parts
35853	Commercial refrigeration equipment
35999	Machinery or parts, nec
37142	Motor vehicle accessories
37147	Motor vehicle body parts
37149	Motor vehicle accessories or parts, nec....
37691	Guided missile or space vehicle parts, nec...
38411	Surgical or medical instruments or apparatus

or jobber). Moreover, for the proprietary deliveries, each member was asked to provide minimum, average and maximum delivery distances. Responses were received from 7 major oil companies operating large proprietary fleets. The responses cover 519 terminals, located in 45 of the 50 states, including Hawaii, plus the District of Columbia [2].

The importance of this supplementary data is shown vividly by Table 6. The seven API companies that responded are responsible for delivering 6.8 million tons of gasoline in the New England region, while both the STS and DES databases show none. Moreover, except for one of the eight regions, these companies deliver more tonnage than shows in the CTS; for example, in the South Atlantic region, their tonnage is larger by a factor of 9.

It may also be important to gather supplementary data on truck-based chemical shipments. OTA asked several major chemical companies to provide flow data regarding their 1983 and 1984 shipments of hazardous materials. Each of them supplied total tons and shipments broken down by origin state, destination state, and commodity. A comparison of these data with the 5-digit STS

TABLE 6

Comparisons of delivered tonnage statistics: API data versus CTS destination state database versus state-to-state database

To region	All statistics in thousands of tons	
	API data for only five companies[a]	CTS 5-digit Dest-State database[b]
New England	6,719	0
Middle Atlantic	17,418	5,020
South Atlantic	18,049	1,836
East North Central	18,536	11,587
East South Central	5,267	5,059
West North Central	3,989	10,652
West South Central	17,832	7,539
All other	20,614	18,791
Total	108,423	60,502

[a]Predominantly gasoline, diesel fuel, and other distillates. The gasoline delivered by these companies represents only about one-third of all gasoline consumption. [b]Gasoline (29111), distillates (29113), and kerosene (29112). The state-to-state database shows only 32,227 thousand tons of gasoline, distillates, and kerosene being delivered by truck. Moreover it shows no petroleum delivery by truck at all in the New England, South Atlantic, West North Central, Pacific Southwest and Alaska and Hawaii regions!

shows that the two sets of data agree on eight of the ten largest movements, but the flows differ in ranking.

While this seems to reinforce the CTS's value, a more detailed analysis indicates that the basis for these rankings may be quite different. When each of the shipper's fifty largest flows (origin state, destination state, five-digit STCC) is checked against the CTS, the highest match rate is only one in three, the lowest is slightly less than one in ten [2].

For rail-based shipments, it may be best to disregard the CTS data. On first inspection, the 5-digit STS database and the corresponding 1977 Waybill Sample appear to agree on the flow volumes. The STS shows 82.4 million tons and 49.8 billion ton-miles versus the Waybill's 72.1 million tons and 43.3 billion ton-miles. The lengths of haul also match almost precisely as well as the breakout of tonnage among chemicals, petroleum, and all other commodities. The databases agree on the three top origin and destination regions, and they agree on seven of the ten top region to region flows. In addition, where one flow is present in the top ten and the other is not, the flow in the opposite database is within the top twenty in all cases.

But the similarities seem to stop there. As Table 7 shows, the databases do not agree on any of the three top commodities in any of the three commodity groups. For chemicals, the Waybill Sample shows sodium alkalies, miscellaneous fertilizer compounds, and anhydrous ammonia, while the CTS shows superphosphate, miscellaneous acyclic organic chemical products, and plastic materials. For petroleum, the databases agree on one of the top three commodities, but the other two are different. For the "all other" category, the databases again agree on only one. More importantly, because the CTS is only for shipments of manufactured goods, it has no data for the top two "other" commodities: crude petroleum and radioactive ores.

For waterborne shipments, the CTS may also be of only limited value. Table 8 shows that the CTS tonnages are only 40% of those in the WCSC database regardless of whether the DES or STS data are used. In part, this is because the CTS does not include crude oil shipments, which is 16% of the total WCSC tonnage. In addition, the CTS shows a different percentage breakdown among chemicals, petroleum products, and other hazardous commodities. For the WCSC these numbers are 9%, 71%, and 21%, respectively; while for the CTS they are 6%, 94%, and 0% based on the STS flows. The two databases agree on two of the top chemical commodities if WCSC 2819 can be considered equivalent to STCC 28182 and WCSC 2810 equivalent to STCC 28122, but the WCSC shows benzene and toluene (2817) as the third top commodity while the CTS shows a aluminum compounds (28196). Finally, the WCSC shows two commodities outside the CTS's purview as being the first and second largest "all other" flows, while the CTS identifies three commodities that do not even fall within the top twenty for the WCSC.

For shipments of hazardous commodities by air, the CTS is the only database available. It may be satisfactory for the shipments from point of manufacture to first destination, but this is not where industry experts say the bulk of the hazardous material shipments are. Rather they are pharmaceuticals (including radioactives) moving between distribution centers, labs, hospitals and doctor's offices.

Supplementary data may be very helpful. This can be illustrated by comparing the CTS with data for a single carrier which was obtained as part of the OTA study. The comparison shows that for 8 of 9 flows originating from a specific region, the STS database shows no flows while the air carrier reports substantial volumes. One of these is the carrier's twelfth largest flow. Of the remaining 72 flows in the region-to-region matrix, there are 11 cases where the carrier's tonnage exceeds that shown in the CTS and three more where the carrier's tonnage is nearly equal to the total shown by the CTS. While the comparison is somewhat unfair because the CTS data is for 1977 and the carrier's data for 1983–1984, it demonstrates that the STS database is not a complete reflection of current airborne traffic flows.

Ways to improve the data

In spite of the deficiencies noted above, the data in the present sources can be used for certain types of safety and policy analyses. As has been shown,

TABLE 7

Principal hazardous commodities by rail: 1977 five-digit CTS vs. 1977 Waybill Sample

WAYBILL	Tons (000)	CTS	Tons (000)	
Commodity		Commodity	STS[a]	DES[b]
Chemicals (total)	42,948	Chemicals (total)	55,422	104,990
28122 Sodium alkalies	4,793	28712 Superphospahte	11,315	17,690
28714 Misc. fert. comp	3,912	28182 Misc. acyc. org. chem. prod.	7,068	8,111
28198 Anhydrous ammonia	3,711	28211 Plast. mat.	7,047	8,979
Petroleum products (total)	19,732	Petroleum products (total)	21,283	30,508
29121 Liq. glass, coal or petrol	6,230	29113 Dist. fuel oil	6,351	10,832
29117 Petroleum residual fuels...	6,064	29121 Liq. gases, coal or pet.	4,389	5,312
29114 Petroleum lub. or similar oils...	2,499	29119 Petrol. ref. prod. nec.	3,099	3,276
All other (total)	9,447	All other (total)	5,723	11,527
13111 crude petroleum	1,735	26111 Pulp	2,596	4,808
10923 Radioactive ores	1,258	20841 Wine, brandy or brandy spirits...	1,064	1,100
20841 Wine, brandy or brandy spirits....	1,047	20999 Food prep. or by-prod., nec.	565	1,161
Total	72,172	Total	82,428	147,025

[a]STS: State-to-state flow data.
[b]DES: Destination state flow data.

TABLE 8

Principal hazardous commodities by water: 1977 five-digit versus 1977 WCSC

WSCS	Tons (000)	CTS	Tons (000)	
Commodity		Commodity	STS[a]	DES[b]
Chemicals (total)	46,337	Chemicals (total)	11,861	21,056
2819 Basic chem, chem prod. nec...	16,597	28182 Misc. acyc. chem. comp...	3,620	3,682
2810 Sodium hydroxide	5,389	28196 Aluminum compounds	1,827	3,291
2817 Benzene and toluene...	4,793	28122 Sodium alkalies	1,572	2,032
Petroleum Products (total	372,871	Petroleum Products (total)	190,153	238,924
2915 Residual fuel oil	136,375	29111 Gaoline	77,595	97,095
2911 Gasoline, inc. add...	96,224	29117 Petrol. resid. fuels	46,938	52,739
2914 Distillate fuel oil	93,379	29113 Distillate fuel oil	43,706	51,567
All other (total)	109,334	All other (total)	704	918
1311 Crude petroleum	86,229	26111 Pulp	607	668
4029 Waste & scrap, nec.	12,022	20841 Wine, brandy, or brandy spiritis...	80	80
2049 Grain mill prod., nec.	5,843	33399 Primary non-ferrous metal ingots...	27	30
Total	528,540	Total	202,754	260,898

[a]STS: State-to-state flow data.
[b]DES: Destination state flow data.

[4,16], region to region flows trip lengths, shipment sizes, and incident rates can be estimated and broken down in a variety of ways. The main limitations are (1) incomplete truck and air data, (2) lack of route and equipment type fields for all but rail and water, and (3) difficult flow selection due to commodity definitions.

Changes should be made to correct these deficiencies. For the Waybill Sample, a consistent, reliable methodology should be developed for selecting the hazardous commodities. Obviously, the 49-series codes should be used; but for earlier years, so should some of the 28- and 29-series codes. For the WCSC data, a set of 49-series type codes should be instituted or else a hazardous materials flag should be added.

The principal challenge, however, lies in improving the CTS data. While nothing can be done about the 1977 data, changes should be made so that CTS databases in the future are more useful from a hazardous materials standpoint. The main goals should be complete coverage, greater commodity specificity, better equipment data, and more route-related information.

To accomplish this, the Bureau should focus on collecting data for truck, air and pipeline. For truck, random samples of "waybills" should be obtained from a representative group of carriers, including private operators. The carriers should be selected from the list maintained by the Bureau of Motor Carrier Safety in its Motor Carrier Census database [10]. Where only paper records are avialable, data encoding will be required, but for most carriers, certainly the larger ones, computerized traffic databases will be available, which means the information can be processed by computer, as is presently done for the Waybill Sample. The only difficulty may be that the carriers will not know or have recorded the route taken. This is a minor deficiency that can be corrected over time. For air, data should also be obtained from the carriers. In this case, however, since the market is dominated by a few firms, the data from these would probably be sufficient. The feasibility of this approach has already been demonstrated in the OTA project [2]. For pipeline, a similar approach should be used since the number of operators is small. For rail and water, comparable data should be developed from the Waybill Sample and the WCSC database, respectively.

But legislative changes may be required to implement this plan. Today, for example, the Bureau cannot approach carriers directly to obtain flow information, instead it must work through manufacturing firms. While this may be a feasible way to collect the data, it severely handicaps or makes impossible the collection of information about the secondary portions of the distribution chain (which are critical pieces for truck and air). The OTA study leaves no question that this information is critical to the development of informed policy decisions relative to the transportation of hazardous materials, and there may be other areas where it is critical as well (for example in the context of general highway and air safety regulation).

Conclusions

This paper has been concerned with the quality of the major databases currently available for analyzing hazardous materials flows. Its purpose has been to illustrate the challenges that presently exist and point to ways in which the databases can be improved. The principal focus has been on three databases: the Waybill Sample, the Waterborne Commerce Statistics Center (WCSC) database, and the Commodity Transportation Survey (CTS), because these form the core of existing commodity flow information.

Generally, it has been found that researchers need to exercise care when using each of these databases to analyze hazardous materials flows. The Waybill Sample has only been using hazardous materials codes since 1976 and the list has changed over time; the WCSC has a very limited commodity specificity; and the CTS has consistency problems and includes only data on shipments of manufactured items from point of manufacture to first destination. Thus in the U.S., a challenge lies ahead to develop more useful hazardous materials flow data.

Acknowledgements

The authors are deeply indebted to OTA and particularly to Ms. Edith Page, for providing the opportunity to perform the work reported here. OTA's counsel and guidance were particularly important in focusing the conduct of the study. Also, the authors would like to thank the numerous people, in both the private and public sectors, who gave of their time and knowledge to add valuable insights.

References

1 Transportation Research Board, Recent advances in hazardous materials transportation research , State of the Art Report 3, Transportation Research Board, Washington, DC, 1986.
2 M.D. Abkowitz and G.F. List, Hazardous materials transportation: commodity flow and incident/accident systems, Final Report, prepared for the Office of Technology Assessment, January 1986.
3 Office of Technology Assessment, Transportation of hazardous materials, OTA-SET-304, Washington, DC, U.S. Government Printing Office, Washington, DC, July 1986.
4 M.D. Abkowitz and G.F. List, A multi-modal assessment of hazardous materials transport risk, presentation before the 66th Annual Meeting of the Transportation Research Board, January 1987.
5 L.L. Philipson, H.S. Napadensky and M.N. Maxey, Hazardous materials transportation risk assessment. In: Transportation of Hazardous Materials: Toward a National Strategy, Vol. 2, TRB Special report 197, 1983, pp. 43–57.
6 K.D. Pijawka, S. Foote and A. Soesilo, Risk assessment of transporting hazardous material: route analysis and hazard management, Transportation Research Record 1020, 1985, pp. 1–5.

7 Association of American Railroads, Rail stations by standard point location codes, Economics and Finance Department, Washington, DC, January 1986.
8 Interstate Commerce Commission, Standard transportation commodity code tariff STCC 6001-M, available in printed form from the Western Trunk Line Committee, Chicago, IL, January 1986.
9 S. Fine and R. Owen, Documentation of the ICC waybill sample, Office of Policy and Analysis, Interstate Commerce Commission, Washington, DC, 1981.
10 G.F. List, M.D. Abkowitz and E. Page, Information sources for hazardous materials flow analyses, Transportation Research Record, 1063, 1985, pp. 15–20.
11 General Services Administration, Code of federal regulations - Transportation, Parts 100–177, U.S. Government Printing Office, November 1984.
12 L. Abbot, G. Bulanowski, B. Foster and J. Jordan, Hazardous materials transportation: A legislator's guide, printed by U.S. Dept. of Transportation National Conference of State Legislatures, Washington, DC, February 1984.
13 ICF, Inc., Economic Analysis of Reportable Quantity Adjustments under Sections 102 and 103 of the Comprehensive Environmental Response, Compensation and Liability Act, Fairfax, VA, March 1985.
14 Army Corps of Engineers, Waterborne commerce of the United States, available from the Waterborne Commerce Statistics Center, New Orleans, LA, 1984.
15 Bureau of the Census, Commodity transportation survey: summary, publication number TC77-CS, available from the Superintendent of Documents, U.S. Government Printing Office, Washington, DC, 1981.
16 G.F. List and M.D. Abkowitz, Estimates of current hazardous material flow patterns, Transp., 40(4) (1986) 483–502.