

**US Army Corps of Engineers**

**Institute for Water Resources**

**LPMS-OMNI and WCSC Data Quality**

**September 2005**

Prepared for  
The Institute for Water Resources

Prepared by  
Mark W. Lisney, Institute for Water Resources, Group R

The government (is) extremely fond of amassing great quantities of statistics. These are raised to the nth degree, the cube roots are extracted, and the results are arranged into elaborate and impressive displays. What must be kept ever in mind, however, is that in every case, the figures are first put down by a village watchman, and he puts down anything he damn well pleases.

**Josiah Stamp**, *Attributed to Sir Josiah Stamp (1849 - 1941) HM  
Collector of Inland Revenue.*

## LPMS-OMNI and WCSC Data Quality

### Purposes

The purposes of this paper are to briefly describe: the major data sources used to conduct Corps of Engineers inland navigation economic analysis; the deficiencies and errors identified in those data sources; and the potential consequences of those deficiencies and errors.

This document was prepared to support the creation of a data analysis and preprocessor module for an IWR Navigation System Simulation model. It is not intended to serve as an enumerator of all Corps databases. Nor is it intended to serve as an encyclopedic history and description of the LPMS, OMNI or WCSC databases.

### Data Set Descriptions

Three databases have been used by the Army Corps of Engineers in economic models as their primary sources of information. Although these databases are also used for purposes other than economic modeling, this paper will focus primarily on economic modeling related issues.

**LPMS-OMNI.** The Lock Performance Monitoring System (LPMS) and Operations and Maintenance of Navigation Information (OMNI) databases are the two systems used by the Corps to collect and store inland navigation lock performance related data. OMNI is currently used in St. Louis, Rock Island, and St. Paul Districts in MVD; Nashville, Louisville, Huntington, and Pittsburgh Districts in LRD; and Little Rock District in SWD. LPMS is used everywhere else. These databases collect and store essentially the same information, but use different user interfaces and storage schemes. The LPMS system is considered the Corps national database. An effort is currently underway to transfer historic OMNI data into the LPMS system. This effort is about two-thirds complete.

The LPMS has been in existence for about 35 years, however most pre-1991 data is unavailable. It is maintained by the Corps Navigation Data Center at the Institute for Water Resources. It started out by using paper forms to record data at the locks. This information was keyed in and stored as ASCII flat files on computer tapes. LPMS moved to web based data entry and Oracle relational database storage in the year 2000. The data items collected has changed twice since LPMS's inception. In 1991, the number of data items collected was reduced and slight modifications were made to data definitions. In 2000, a major change was made to the way data is recorded and again, data definitions were modified. In 2000, LPMS moved from a lockage based approach to a cut based approach.

The OMNI system was developed by Rock Island District and was first run in 1980 on a Burrough's mini-computer. It was converted to Windows and Oracle for use by MVD in 1995. LRD started using the OMNI system in late 1997.

LPMS and OMNI serve as the only sources of data regarding lockage times, lockage types, and vessel counts of tows, recreation craft, passenger vessels and other vessel types. They are the timeliest sources of barge counts, tonnages by general commodity code, and lock outages. LPMS-OMNI is considered the most accurate source for lock barge counts.

**WCSC.** Waterborne Commerce Statistics Center (WCSC) data is submitted to the Corps by towboat operators, barge operators, and cargo manifests and custom clearing for foreign data. It contains information about the amount and types of equipment using the waterway system, how the equipment moves around the system, and the types and amount of commodities moved by the equipment. The collection, error checking, and data storage is managed by the Waterborne Commerce Statistics Center in New Orleans, LA. The entire WCSC database covers foreign and domestic deepwater data, but this document addresses only the inland waterway portion of the data.

The commodity descriptors are much more detailed in OMNI than those collected by LPMS-OMNI. It has always been assumed that tonnage loadings per barge are more accurate in WCSC than LPMS-OMNI. Our economic modeling efforts use WCSC data for estimating loadings per barge, estimating historic origin-destination-commodity flows, predicting future origin-destination-commodity flows, and estimating economic models. WCSC data is also used for budget preparations, state-to-state and other area-to-area commodity flow reports, and waterway commerce reports.

### **Data Inaccuracies and Deficiencies**

Nearly a dozen people familiar with these data sources were consulted during preparation of this paper. The remainder of this document lists and describes the issues raised by these people. Although this may not be an exhaustive recitation of all the limitations and deficiencies one may find, the most important issues are likely presented.

### **LPMS-OMNI**

**Deficiencies.** This section describes the information that the queried sources would like to see, but which is not included in OMNI or LPMS.

***Lightboats locking with tows:*** If a lightboat locks with a tow, the lockage record will record the fact that x number of lightboats locked with the tow, but the vessel number of the lightboat(s) is not recorded. (The vessel number is recorded in OMNI, but it is not transferred to LPMS). Several people cited this as a shortcoming in the data. One powerful use of the data is to track tows. Where they begin, change configuration, and end. This task becomes more difficult if a towboat travels through one or more locks as a lightboat. The data contain movements that appear to be impossible. For example, a towboat makes two successive upbound movements through a lock without an intervening downbound movement. By not recording every lockage of a towboat, the task of

following towboat movements becomes more difficult, and the analyst is left to assume that seemingly impossible moves are caused by the lack of lightboats being recorded with tows. Pre-1991 LPMS had fields for recording these light boat vessel numbers, but most users were not entering the data. A 1991 LPMS task force decided this field was serving no purpose. From 1991 to present, the vessel numbers of light boats locking with tows does not appear in the data.

***Data describing conditions during lockages:*** Several people requested that data be included with the lockage record that better describes river conditions. For example, was a lockage made during the day or at night? What were the flow conditions in the river? What were the upstream and downstream gage readings? Prior to 1991, there were several fields available to record gage readings, wind conditions, river flow conditions, and several other items. The 1991 task force eliminated these fields. I'm not sure how much of this data is available in OMBIL or other databases maintained by the Corps. When I was a member of the LPMS task force in 1999 we decided that we didn't want to record the same data in two places. If this is available somewhere else, then it doesn't necessarily need to be available in LPMS-OMNI.

***Missing vessel numbers:*** One person indicated that of 70,000 flotillas analyzed, he was unable to match the recorded vessel number with the vessel number database for 5,500 flotillas. It is impossible to know whether this was caused by improperly entered vessel numbers, or a deficiency in the vessel number database. Incorrect vessel numbers are a difficult problem to catch programmatically. Typical errors are reversing digits, and entering the wrong vessel number. A lot of vessels don't have Coast Guard assigned vessel numbers and vessels are being bought and sold all the time. It usually takes a person to sort out vessel errors. One person responsible for the current LPMS system indicates that this should not be a problem with 2000 to present LPMS data. The LPMS system will not accept vessel numbers that are not in the Vessel Table of the LPMS system. LPMS provides a means to enter new vessel numbers into the Vessel Table. These new numbers are then checked against the Coast Guard database.

***Impaired performance:*** One person responded that more data should be available in each lockage record regarding periods of impaired performance and the cause of the impaired performance.

***Twin Datasets:*** One person stated the splitting of responsibilities and data management is a major shortcoming of ACE data collection efforts.

**Data Inaccuracies.** This section describes errors identified in the OMNI or LPMS databases. Errors discovered in WCSC data will be presented in a separate section. This section is divided into those errors that could likely be caught with error checking routines, and those that likely wouldn't be caught with error checking.

**Catchable**

**Component Lockage Times:** Examples of these types of errors include:

1. turnback approach starts at the same time as the previous end of lockage.
2. the same vessel number is at two locks at the same time
3. fly approaches begin later than the arrival time.
4. turnback approach starts before the previous end of lockage
5. the dimensions of a single cut flotilla do not fit into the chamber. Either the number of cuts is wrong or the flotilla dimensions are wrong
6. the variable that indicates the number of cuts in a flotilla doesn't agree with the number of unique start of lockage times for that flotilla
7. end of lockage occurs before start of lockage (I didn't see this in LRD or Upper Miss data)
8. extremely long lockage times (3 days) without an accompanying stoppage record

### **Not easily Catchable**

1. use of a standard tonnage value instead of a more accurate estimate. The analyst making this comment recognizes that recording more accurate tonnages in LPMS is very difficult if not impossible.
2. rounding lockage times to the nearest 5 minutes. During the capacity analysis for the Ohio River Mainstem Study, it was found that several locks round all times to the nearest 5 minutes (11:45 instead of 11:43 or 11:47). The data rounding was so bad at several locks that the times couldn't be used. One of the most important pieces of information collected in LPMS-OMNI is the processing time data. If they are improperly recorded, most of database is serving little purpose and is simply a great waste of time.

### **Inconsistent-Improper Recording Standards**

This section contains examples of instances where data are not recorded consistently across District, Divisions, or rivers, or where data is not recorded in accordance with the LPMS/OMNI definitions.

1. Apparently it is common practice on the Mississippi River for a tow to call in and receive a place in queue long before the tow actually arrives. The following is from one of the people queried:  
“We heard anecdotal evidence from lockmasters and pilots that it is common practice on the Upper Miss to call in early when there is a queue so that everybody preserves their spot in line. We also heard anecdotal evidence from the same groups that the tows then decrease their speeds when they are approaching an existing queue to better time their arrival and conserve fuel. Given the way the arrival time is generated in LPMS it was impossible for us to measure the slower speeds and possible fuel savings.”
2. Inconsistent interpretations of what constitutes a stoppage.

This issue has been around since the beginning of LPMS. While economists use this data to determine the availability of a chamber for capacity calculations, Operations personnel have used this data for personal performance appraisals. There was a bias years ago to NOT record stoppages because it reflected negatively on the lock personnel. Consequently, very few stoppages were recorded.

On other occasions, lock personnel were under the impression that a stoppage shouldn't be recorded unless it exceeds a certain duration, say 1 hour or 8 hours.

Another example is events like fog were not recorded because "the lock is available, we can't help it no one wants to approach in the fog."

Similarly, an entire winter shutdown may not be recorded because the lock could serve a vessel, it's just that no one is there because the river is froze over.

One source found many closures which appeared in other sources did not appear in LPMS-OMNI. He concluded that if a lock has two chambers, and one chamber is serving all traffic, the other chamber is not marked as closed, even if it is closed. It's like if the lock can serve traffic, then no closure occurs.

One source questioned whether closures that don't affect traffic are recorded consistently from lock to lock or District to District.

3. Inconsistent interpretation of what constitutes a "Scheduled" or "Unscheduled" stoppage.

Some locks record a stoppage as being scheduled if a Notice to Navigation has been issued for that stoppage. Others record a stoppage as scheduled only if it was part of a scheduled maintenance program.

For example, the McAlpine closure in 2004, which was highly coordinated with shippers and carriers, and had several NTN issued, was recorded as Unscheduled. From an economics standpoint, this closure should be treated as scheduled. From an engineering standpoint, the closure could be considered unscheduled because it was caused by an unexpected failure which was discovered via dive inspections.

Likewise, Markland L&D main chamber has had a dewatering inspection closure each of the last few years. This closure occurs because the miter gates are in very poor condition and must be inspected and repaired every year. In spite of the fact that NTN are issued each year, these closures are recorded as unscheduled. Apparently the closure is recorded as unscheduled because, under normal maintenance practices, this chamber would be closed only once every 5-7 years.

It appears that the data field indicating whether a closure is scheduled or unscheduled serves two purposes. Some closures can be correctly recorded either way, depending on the purpose.

#### 4. Inconsistent databases

One source noted that these are live databases. If errors are found in the data, the original data is corrected. This leads to data sets which may be different, depending on when they were downloaded from the master source.

#### 5. Gaming the System

Two sources noted they found some evidence of “Gaming the System”. By this they mean people are recording false data to serve their own purposes, or to serve what they perceive to be the purposes of others. For example, one lock is shown as being busy from midnight until 1 AM 98% of the time. This occurs at a lock where utilization is only 40% in the previous and successive hours. Further inquiry found no legitimate reason to explain this vast utilization discrepancy. Another example is evidence where extra records are added at the end of a shift, or lockages are not recorded.

### **WCSC Data**

Historically, the Corps has considered WCSC data as the most accurate source for tons per barge estimates. There is no viable way of cross-checking tons/barge figures with LPMS-OMNI because it is well known that LPMS-OMNI data are usually rounded to 1500 tons/barge for all jumbo size barges. Therefore, we are left to assume that tons/barge data in WCSC is correct.

Regarding barge counts and trip counts, we have considered LPMS-OMNI as the most accurate data source. Lock operators may have a difficult time determining the amount of commodity, or commodity type, in a barge. However, it is quite easy to count the number of barges in a tow, and record the number of tows that traveled through a lock during a shift, or day. When we compare LPMS-OMNI barge trips with those recorded in WCSC, we find that many trips are not recorded in WCSC. This is especially a problem in the post 1999 data.

The accuracy of WCSC data has come under question during the last 6 months. Performance based budgeting prompted the TVA to closely look at WCSC data. They found that trips on the Tennessee River were underreported by as much as 15%. This underreporting led to a recommended cessation of FY06 funding for the Kentucky Lock expansion project.

Nationwide, comparison of WCSC with LPMS data for 2003 produced a 34% unmatched ratio. This means that given a specific vessel found on a lock report, there is about a 2

out of 3 chance of finding that exact vessel in the WCSC database for that lock in the same time frame.

Another comparison of WCSC and LPMS data found barge trips were unreported by 8.2% for the nation as a whole. The underreporting showed a large variation by river system, varying from as much as 48% on Freshwater Bayou to as little as 1.5 % on the Monongahela River.

Other problems found in WCSC data include computational logic errors in the data base, out of date data files and tables, data estimation, carrier under reporting, carrier gaming to avoid fuel taxes, data input errors in route codes, default values, and location codes. Some evidence of gaming, where towboat operators submit trip reports to reflect fuel taxes paid, are evident in WCSC.

### **Consequences**

Data deficiencies will lead to either, an analyst searching for the information in other sources, or replacing recorded data with educated assumptions. For example, if towboats cannot logically be tracked because lightboat vessel numbers are not recorded, the analyst must either try to track the boat in WCSC data, or assume that the illogical movement is correct.

Inaccurate data force data analysts and users to spend additional time and resources devising ways to capture and discard the errors. For example, component lockage time errors complicate the process for developing lockage time input to economic models. Any data analyzer or preprocessor which attempts to summarize input data is going to have to check for data errors before the data is summarized as input. Since there are so many different ways in which the data can be in error, this process is going to be unnecessarily complex and expensive.

Improperly recorded data, such as recording a vessel as having arrived when it hasn't, leads to inaccurate reporting statistics.

Inconsistent and nonexistent stoppage data has far reaching consequences, not just modeling consequences. When Corps HQ put out a nationwide data call last year for information related to unscheduled closures, the inconsistent interpretation of what constitutes a scheduled or unscheduled closure has direct bearing on the results of the data call. The economic consequences of scheduled closures are much less than those of unscheduled closures. If the 2004 closure of McAlpine had been unscheduled from an economic standpoint, it's likely that several businesses would have closed and never reopened. Many shippers expressed the need for a long lead time to build inventories before the closure.

When an analyst is ready to develop downtime files for the NaSS program they are going to have two choices. Either they will develop downtime events from LPMS-OMNI data and use the results without modification or they will develop downtime events from

LPMS-OMNI and then go through a process of review and analysis before using the data. Clearly it would be less expensive if an analyst could take LPMS-OMNI downtimes at face value.

Another consequence of improperly recorded data is the effect it will have on being able to validate reliability models. A great deal of effort is going into the development of reliability models which will be used to predict future component failures. How will those models be validated if the data they are validated against is inaccurate?

One consequence of inaccurately recorded WCSC data has already occurred. As stated above, the initial recommendation was to discontinue funding for the Kentucky Lock Improvement project in FY06. This recommendation was attributable to inaccurately recorded barge trips in WCSC data. The FY06 Energy and Water Appropriation bill has not been passed as of this date. However, indications are that partial funding has been restored. Another logical consequence of underreporting in WCSC is the Inland Waterway Trust Fund is being shortchanged.

From a computer modeling standpoint, it puts the analyst in a difficult position if he has to choose between believing a known inaccurate LPMS-OMNI database and a known inaccurate WCSC database.

### **Summary Thoughts**

First, it is worth noting that the explicit purpose of this paper is to uncover and document errors and deficiencies in LPMS, OMNI, and WCSC databases. The reader should not project from this paper that all LPMS-OMNI-WCSC data are bad. Indeed, most inland waterway data are recorded properly. Analysts who have used data from other transportation modes indicate that inland navigation data are far better than rail or truck. They say we should celebrate the fact that our data are so good. This does not mean improvements are unattainable.

In one of the WCSC emails, where data accuracy was being discussed, the following statement was made:

“The choice in the future is to either continue beating up on the inland towboat and barge operators to gain their accurate data submission OR turn to a new data collection alternative using electronic methods....”

For LPMS-OMNI, one could modify that statement to say:

“The choice in the future is to either continue beating up on the Lock Masters and Lock Operators to gain their accurate data submission OR turn to a new data collection alternative using electronic methods....”

I'm not sure if electronic recording methods will eliminate all problems, but it deserves serious consideration.