FINAL REPORT

AN ANALYSIS OF U.S. TANKER AND OFFSHORE PETROLEUM PRODUCTION OIL SPILLAGE THROUGH 1975

Robert J. Stewart
Michael B. Kennedy

Martingale, Inc.
Cambridge, Massachusetts

Report to

Office of Policy Analysis
U.S. Department of the Interior
Washington, D.C.

Contract No. 14-01-0001-2193

10 May 1978

ACKNOWLEDGMENTS

The work and dedication of three individuals were essential to this report. H. S. Lahman's assistance was invaluable in combining poorly documented computer tapes and badly formatted printed tables into a well-formatted, accessible computer database. J. W. Devanney provided the critical review. Many of the concepts presented in this report owe their genesis or format to him. We would also like to thank Beth Parkhurst for her help in the editing and production of this report.

EXECUTIVE SUMMARY

This is the final report for Contract #14-01-0001-2193. It is the fourth and last item provided for in this contract, and, in the context of the other three items, it completes the analyses required for the revision and updating of the Offshore Risk Analysis Group's spill incidence and volume algorithms. These algorithms provide techniques for estimating the number of oil spills and the volume of oil spilled for impact analyses of prospective offshore petroleum developments. The Offshore Risk Analysis Group's previous algorithms relied on data taken prior to 1973. The present works incorporate data through 1975. The present study also incorporates a novel volume distribution methodology.

Four distinct spill sources were considered in these studies: tankers, subsea pipelines, production platforms, and single-buoy mooring (SBM) offshore loading facilities. An intensive examination of the available data for these spill sources was undertaken in our previous contract (Stewart, 1976). It was concluded in that study that all but the SBM could be analyzed using U.S. data, although care had to be taken even with the U.S. data due to erroneous source classifications.

As the first item under this contract, a search was made to find North Sea SBM data to supplement existing U.S. data resources. United Kingdom officials and oil spill authorities were contacted by phone and letter with the result that Mr. A. D. Read of the Petroleum Engineering Division, Department of Energy, London, was finally identified as the custodian of the requisite data. Unfortunately, the North Sea data was found to be confidential, and so no further progress was possible in this area. Our report of 22 May 1977 summarized these inquiries and appended all related correspondence.

The second item in this contract was the transfer of our computer programs to the Offshore Risk Analysis Group. This was initially envisaged to include primarily the gamma family posterior volume distribution. However, in the course of performing the analyses for this study, we found it necessary to correct the existing gamma-based posterior; to develop new posterior distributions for the lognormal and inverse-gamma families; and to develop a Bayesian hypothesis test to select the appropriate posterior. All requisite programs for these analyses have been delivered to the Offshore Risk Analysis Group. Appendix C of this report is a draft of a paper outlining this novel volume methodology.

The third item in this contract was an analysis of oil spill risk under extreme environmental conditions. General techniques were developed to allow simulation of the oil spill history of offshore developments subject to hurricanes and earthquakes. This work is summarized in our report of 6 June 1977, "A Monte Carlo Platform Failure/Oil Spill Model." Again, all pertinent programming has been delivered to the Offshore Risk Analysis Group.

This report, the fourth item in the contract, deals with spillage from tankers, production platforms, and subsea pipelines. The primary data sources were the USCG Pollution Incidence Reporting System (PIRS) and the USCG Event File. Supporting information was obtained from Martingale's Master Vessel File (MVF), the American Bureau of Shipping Register, the USGS Platform File, and numerous other files maintained by the Corps of Engineers, the Maritime Administration, the Coast Guard, and the Census Bureau.

The data was such that the analysis of U.S. tanker spillage was the most definitive of those attempted. Because most U.S. tankers ply between U.S. ports, nearly all of their spills were known, and this information was nicely supplemented with yearly activity figures from MarAd and vessel characteristics from the MVF. Consequently, we found very useful ways of describing the propensity of U.S. tankers to have an oil spill, and we found similarly definitive spill volume distributions. We found that old tankers have higher spill rates. We also found that other ways of classifying spills were not so useful; specifically, size and, by proxy, port calls, were not related to spill number. The analysis revealed that a typical U.S. tanker has a spill rate of about .4 spills per year. An approximate analysis was also conducted for foreign tankers, but data problems limit the reliability of this analysis.

The analysis of offshore platforms revealed that either platform-years or annual production were reasonable exposure variables for a spill incidence model. We developed our final model using platform-years as the exposure variable, and found, rather surprisingly, that old platforms have fewer spills on a unit basis than new platforms. New platforms appear to be subject to a run-in period, during which they are prone to experience three or four spills in a one-year period.

The subsea pipeline analysis was done on the basis of pipeline-mile-years as the exposure variable. Unlike the tanker and production platform analyses, which relied on quantitative statistical tests for their validation, the pipeline model relies on assumptions for support of its incidence model. In this sense, it is the least satisfactory of the results derived in the study, but there was no practical alternative to this assumptive approach, due to the nature of the pipeline data.

In addition to these products, which are directly applicable to the Offshore Risk Analysis Group's objectives, we also learned a great deal about the various federal spill data sources. We summarized the difficulties we encountered with these files and made a number of recommendations for improving their reliability and usefulness in future studies of this type.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Introduction

The environmental consequences of developing petroleum resources on the outer continental shelf (OCS) are perceived as the principal disbenefits of the Department of the Interior's OCS leasing program. While such developments can affect the environment in a variety of ways, it now appears that the central issue is the degradation of the environment that might be caused by the oil spillage that accompanies such activities. This report examines newly available data and develops new quantitative techniques for application to the oil spillage problem. These results are to be applied in predicting the nature of the oil spillage problem for candidate OCS developments. Not addressed in this report are the problems of where the oil will go and how it will affect the environment.

It is now generally accepted that the spillage problem cannot be properly quantified with a few simple averages. The potential impact of the Argo Merchant oil spill, for example, is not discernible in the statement ".006% of all the oil brought into U.S. harbors in 1976 was spilled."* The technique we adopt to highlight the disproportionate importance of the rare large spill is to analyze the problem in terms

---

*This figure is based on the volume of oil spilled by the Argo Merchant and the volume of oil handled in U.S. waters by tanker in 1975.

of spill frequency models and spill volume distributions. The former are used to predict the number of events that might occur; the latter, to predict the volume spilled given the event. The predictions are made in terms of probabilities, that is, "n spills will occur with probability p(n), and less than x gallons will be spilled with probability p(x)." This information can be used to generate simple averages if desired, but a more typical application might involve comparing several alternatives based on their probability of experiencing one or more very large spills.

Data on spill incidents is readily acquired from a variety of sources. The U.S. Coast Guard maintains a Pollution Incident Reporting System (PIRS). This data presumably includes all spills in U.S. waters within the three-mile limit and those spills over 50 barrels (BBL) from U.S.-controlled sources outside this limit. The U.S. Geological Survey maintains an accident file that includes all spills over one barrel (1 BBL = 42 gallons) from spill sources operating on U.S. leases. Certain states (Louisiana, Texas, California, and Maine, to name a few) maintain spill statistics relating to petroleum production or transportation activities within the state. Spill data from tankers on a worldwide basis is available from Lloyd's Weekly Casualty List. A number of private firms and the U.S. Coast Guard have computerized summaries of the Lloyd's data to facilitate access.

In view of this apparent wealth of historical information it might seem paradoxical that our ability to make spillage predictions is still rather rudimentary. However, these spillage records often do not include the information required to develop frequency-of-occurrence models. Such ancillary data is referred to in the recent literature as "exposure data". This data includes such descriptive information as the age of the spill source, the number of active spill sources in an age class, and the volume of oil handled annually by the spill source. Parameters such as these frequently form the basis for predictive models of spill incidence. In the past, such information was not collected. With today's data, we can estimate the required values to a fair degree of accuracy, although extensive cross-referencing is needed.

The analysis of spill volume distribution was also complicated in the past by the lack of suitable ancillary data regarding the specifics of a spill incident. An analogy can best illustrate this point. Suppose we had a collection of apples and watermelons and we were asked to describe the weight of any one member drawn at random. Our task would be simpler and the resulting prediction more accurate if we were allowed to determine first the type of fruit drawn. In the same way, when we deal with all spills occurring from ships, we can expect some reduction in our prediction's uncertainty if we can first ascertain certain key features of the incident such as whether the hull was ruptured.

This report examines the PIRS data for 1973-1975 and the USGS data for 1971-1975 with these problems in mind. Modifications to the information encoded in PIRS in 1973 and the development of several new files have allowed us to develop suitable exposure data. The focus is strictly on those systems used in the production and transportation of crude oil from OCS regions. Our scope is further limited to those systems for which there is an adequate historical base. Not included for this reason are production schemes based upon subsea completion techniques. The methodology for the spill volume model is an extension and substantive correction of the Bayesian techniques used by Devanney and Stewart (1973). The spill frequency model is developed along more classical lines.

Background

A variety of analogous studies have been made in the past five years. Devanney et al. (1972) attempted to apply the 1970-1971 PIRS data to spillage predictions for hypothetical petroleum development on New England's Georges Bank. This analysis was refined and the data base expanded to include the 1972 PIRS data, worldwide ship spillage data based on ECO Inc. records, and specific state spillage records based on EPA data in a sequel by Devanney and Stewart (1973) for the Council on Environmental Quality (CEQ). This report was again directed toward offshore petroleum development spillage. Paulson, Schumaker and

Wallace (1973) applied Stable law distributions* to the 1970-1972 PIRS data from the Eighth and Thirteenth Coast Guard Districts. The problem they attacked most successfully was estimation of the total volume of oil spilled in a number, k, of incidents (k sums). It is difficult to see how this information can be interpreted and implemented for contingency planning for a particular event, but it is apparent that such an approach has several applications to more general questions. In particular, the Stable law assumption doesn't care about the apple and watermelon content of the sample (see above) and so it might be applied to all spills of any one source class to yield estimates on the total spillage from the source given estimates of the number of spills. As these authors point out, this information can then be used in allocating monitoring and supervisory resources. Such an approach would also be useful in establishing total, lifetime spillage figures for an offshore development, although we have not pursued the matter here. Umlauf, Pizzo and Huster (1974) analyzed Standard Oil of California conventional buoy spillage records for 1968 through 1971 and U.S. Coast Guard Captain of the Port quarterly casualty records. Their purpose was to estimate the spillage that might accompany petroleum transfer operations in the State of Washington.

---

*Stable law distributions are a class of continuous univariate probability density functions characterized by their property when summed of retaining their initial form. The Normal distribution is one example of this class of function, as is the Cauchy.

They found a strong correlation between vessel groundings and collisions with number of vessel port calls, for harbors in several different regions. They then estimated the fraction of these events that would result in oil pollution. Swift (1973) looked at spillage records from the Cook Inlet region and derived average spillage figures for use by the U.S. Coast Guard in anticipating the nature of the pollution problems that might arise in Alaska in the 1980s. Most recently, Bayer and Painter (1977) presented a variety of results relating to oil spillage from offshore developments, as well as onshore pipelines. They discussed the problem of establishing suitable exposure parameters for the spill incidence models. Concluding that there was no good way to handle the problem with the data they were using, they simply listed a number of incidence parameters. They plotted spill volumes on lognormal probability paper and proposed that the lognormal form be used for spill volumes. They showed a useful correlation between mean spill size and onshore pipeline diameter. They also demonstrated that vessel age was not a useful predictor of spill incidence or size. The data bases they worked with included worldwide vessel casualty data, Canadian and European onshore pipeline data, and USGS offshore spillage data.

Methodology

Because of the difficulty in obtaining suitable exposure data, the spill incidence models in the studies referenced above relied heavily upon assumptions. Spill volume models have tended to be rather assumptive due to the theoretical difficulties that accompany the analysis of strongly non-normal probability distributions. These characteristics are undesirable. Our general goal in this study was, therefore, to develop new models that relied more on the available data and less on assumptive structure.

Beginning in 1973, the data collected by the Coast Guard PIRS was modified to include a variety of interesting exposure information for each recorded spill incident. It was apparent, for example, that some analysis of spill incidence and environmental factors could be made based solely on the PIRS wind velocity, wave height and current data. Further, the inclusion of source identifier information in the revised data opened up the possibility of extensive cross-referencing with other data bases.

A substantial effort was made to identify these other data bases. The USGS event, structures, LPR10, and pipeline files were identified and acquired, and then cross-referenced where possible. Our preliminary results at this task are reported in our earlier report, Stewart (1976) and also summarized in Stewart (1977). The following section and Appendix A document our final results.

Following the cross-referencing, we attempted to ascertain the quality of the data by intercomparison. Our general conclusion was that the data was not as reliable as we had originally hoped, although it was sufficiently improved to allow substantive improvements in spill incidence models. To avoid hiding the data deficiencies in a haze of analytical techniques, we have adopted the philosophy for this report that simple is better. This is reflected in our decision to stick to one-dimensional exposure models. That is, rather than analyze the data from the standpoint of multiple regression and factor analysis, we consider here only simple regressions. To make a more sophisticated approach credible in view of the data problems was beyond our resources. In any event, we found some useful predictors so little was lost.

The desirability of modifying the spill volume methodology used in Devanney and Stewart (1973) didn't become apparent to us until we had first corrected some errors and inconsistencies in our previous programming and theory. However, we then found that the Gamma family, despite its versatility, wasn't producing very good fits to the observed spill volume data. We then re-examined the whole volume methodology, and concluded that other families of continuous, unimodal univariate probability density functions (PDFs) should be included. Retention of the Bayesian framework was judged desirable because it seemed likely that we would still have the small-sample problem that so befuddles classical statisticians.

On practical grounds, we limited the selection of possible PDFs to the two-parameter Gamma, Lognormal, and Inverse Gamma families. These families exhibit increasingly strong tails, a property that is most evident in the relationship of the kurtosis to skewness (provided that these parameters are not infinite). The Gamma has the smallest kurtosis for a specified skewness of the three PDFs, while the Inverse Gamma has the largest. The details of the derivations and programming for the posterior volume distributions may be found in Stewart and Kennedy (in draft). Copies of the programs BAYLOG and POSTVOL and subroutines ASYSAN and QADSAN have been supplied to the Offshore Risk Analysis Group of the Department of the Interior. These programs calculate the posterior volume distributions for the Lognormal (BAYLOG) and Gamma and Inverse Gamma (POSTVOL) PDFs given the sample's sufficient statistics.

We were then faced with the problem of choosing between the three candidate PDFs. Classical goodness-of-fit tests were considered, with the idea of comparing the posterior cumulative distributions with the observed histogram. This approach, however, short-circuits the whole Bayesian methodology and has substantial conceptual difficulties. We finally hit upon a Bayesian hypothesis test of the underlying distribution based on posterior likelihood functions. The theory underlying this method is analogous to that used in Bayesian econometrics to weight candidate regressions in multiple-regression analysis (c.f. Zellner, 1971). However, we are unaware of an equivalent result in our application. Details of the method may be found in Stewart, Devanney, and Kennedy (Appendix C). Again, the

Offshore Risk Analysis Group was provided with the pertinent program (POSTLIK).

Although the details of the method are best left to the papers referenced above, it is useful to list the pertinent formulas underlying the technique. These consist of the underlying probability density functions, the specifications of the sufficient statistics, the prior distributions assumed for the parameters underlying the PDF, the posterior volume distribution, and the posterior likelihood function. These formulas are shown in Table 1. The prior distributions on the underlying parameters were selected on the basis of Jeffrey's recommendations for parameters with semi-infinite and infinite ranges. They represent complete ignorance of these parameters prior to looking at the sample (we considered them to be noninformative priors). See Jeffreys (1967), Zellner (1971), and Lindsey (1970) for the theory underlying Bayesian analysis in general.

Summarizing, the spill incidence models were developed along simple classical grounds, with no strong overlying methodology. We simply tested those exposure parameters in which we had some confidence. The spill volume models were developed with a more rigorous Bayesian methodology that was structured on the assumption that spill volumes were distributed according to one of three postulated PDFs (i.e., the two-parameter Gamma, Lognormal, and Inverse Gamma). In this sense, the present study might

TABLE 1
FORMULAS UNDERLYING THE SPILL VOLUME METHODOLOGY

Name	PDF	Sufficient Statistic	Prior on Underlying Parameters	Posterior Volume PDF	Posterior Likelihood Function
Gamma	$\frac{\lambda(\lambda v)^{r-1}e^{-\lambda v}}{\Gamma(r)}$	n $\sum_{i=1}^{n} v_i \equiv S_r$ $\prod_{i=1}^{n} v_i \equiv P$	$\frac{dr\,d\lambda}{r\lambda}$	$\frac{1}{v}\,\frac{n}{n+1}\left(\frac{n-1}{a}\right)\frac{S(a(v),n+1)}{S(a,n)}\left[\frac{a}{a(v)}\right]^{n}$	$\frac{1}{2}\left(\frac{a}{\zeta^2}\right)^{n-1}\frac{1}{p\zeta}\,\frac{\Gamma(2n-1)}{\Gamma(n-1)}\,\frac{S(\zeta,2n)}{S(a,n)}$
Lognormal	$\frac{e^{-\frac{(\ln v-\theta_1)^2}{2\theta_2}}}{v(2\pi\theta_2)^{1/2}}$	n $\sum_{i=1}^{n} \ln v_i \equiv \bar{v}\cdot n$ $\sum_{i=1}^{n} (\ln v_i - \bar{v})^2 \equiv \tilde{S}\cdot(n-1)$	$\frac{d\theta_1 d\theta_2}{\theta_2}$	$\frac{1}{v}\left(\frac{n}{n+1}\right)^{1/2}\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\,\frac{\left[1+\left(\frac{n}{n+1}\right)\frac{(\ln v-\bar{v})^2}{(n-1)S^2}\right]^{-\frac{n}{2}}}{(\pi)^{1/2}(n-1)^{1/2}}$	$\frac{\Gamma(n-1/2)\left[\prod_{i=1}^{n} v_i\right]^{-1}}{\Gamma\left(\frac{n-1}{2}\right)(2\pi)^{n}[\pi(n-1)]^{\frac{n}{2}}}$
Inverse Gamma	$\frac{\lambda^{r}(v)^{-(r+1)}e^{-\frac{\lambda}{v}}}{\Gamma(r)}$	n $\sum_{i=1}^{n} \frac{1}{v_i} \equiv \Sigma$ $\prod_{i=1}^{n} v_i \equiv P$	$\frac{dr\,d\lambda}{r\lambda}$	$\frac{1}{v}\,\frac{n}{n+1}\left(\frac{n-1}{b}\right)\frac{S(b(v),n+1)}{S(b,n)}\left[\frac{b}{b(v)}\right]^{n}$	$\frac{1}{2}\left(\frac{b}{\zeta^2}\right)^{n-1}\frac{1}{p\zeta}\,\frac{\Gamma(2n-1)}{\Gamma(n-1)}\,\frac{S(\zeta,2n)}{S(b,n)}$

where:

n = number of samples

$\underline{v} = (v_1, v_2, \ldots v_n)$, the sample vector

$a = -\ln\left(\frac{P}{S_r^n}\right)$

$a(v) = a + (n+1)\ln\frac{(S+v)}{S} + \ln\frac{S}{v}$

$\zeta = 2n\ln(2) + 2a$

$b = \; n(p\Sigma^{n})$

$b(v) = b + \ln(1 + v\Sigma) + n\ln\left[\frac{1+v\Sigma}{v\Sigma}\right]$

$\zeta = 2n\ln(2) + 2b$

$S(x,m) = \frac{1}{\Gamma(m-1)}\int_0^{\infty} e^{-\beta}\beta^{m-2}\,\frac{\Gamma\left(1+\frac{m}{x}\beta\right)}{\Gamma^{m}\left(1+\frac{\beta}{x}\right)}\,d\beta$

still be considered assumptive. Practically, however, it represents a significant step towards letting the data speak for itself.

Data sources

The impetus for the study preceding this report, Stewart (1976), came largely from the expectation that suitable exposure data could be obtained for the tanker, pipeline, and production platform spillage based on the revised PIRS data, the Martingale Inc. Master Vessel File (MVF), and the U.S. Geological Survey's LPR10 offshore production file. In our previous report, Stewart (1976), it was shown that this expectation was not completely fulfilled. In particular, the PIRS data was found to be unsuitable for distinguishing offshore production from offshore pipeline spillage. Three out of four common carrier pipeline spills over 50 BBL in the period 1973-1975, for example, were labelled production platform spills in the PIRS data. In the study at hand, our major conclusions for the pipeline and production platform categories are therefore based on the U.S. Geological Survey Event file. Several listings and versions of the file were obtained from Mr. Elmer P. Danenberger of the USGS Conservation Division in Reston, Virginia, and from Mr. Doug McIntosh of the USGS District Office in Metairie, Louisiana.

The PIRS data is nevertheless of central importance to this study, because it is our sole data source for tankers, and we have used it to investigate the possible effects of adverse weather on spill incidence from the combined pipeline and platform sources. The PIRS data uses a number of codes to describe a spill incident. The spill source, for example, is characterized by a three-digit source code and an eight-digit identifier. The source code for tankers falls in the range 010-019, depending on the vessel's GRT, and the vessel is identified either by its radio call sign or its registration number. Offshore platforms are identified by the source code 506 and an identifier based on a list of companies. The volume spilled and the volume recovered are recorded in gallons and the material spilled is identified by a four-digit code. Crude oil has the code 1000 or 1001, depending on whether it is a "light" or "heavy" crude. The location of the spill is nominally coded by either the latitude and longitude to the nearest minute, or by the nearest mileage marker for navigable inland rivers. However, some offshore spill incident locations are coded by area and block number. The cause of the spill is coded using a two-letter code, and the nature of the operation in progress at the time of the spill is listed according to a two-digit numerical code. There are also numerous data fields relating to clean-up expenses and legal action taken. All in all, a PIRS record consists of 435 coded characters, although about one-third of these are at present used for nothing more than spacers.

Theoretically, the use of codes offers many advantages in data reduction over written narratives, since the coding provides a uniform method for classifying the spill incident.

However, the coding language must be carefully designed if it is to achieve this purpose. Specifically, it must provide the encoder with an exhaustive and non-overlapping set of choices. The PIRS code, unfortunately, does not do this for several of the data fields. The identifier code for all non-ship sources, for example, provides fewer than 1,000 real coding choices, and no installation is uniquely identified. Thus, this field is not exhaustive. Alternatively, the cause and operation codes are a hopeless muddle of overlapping and ambiguously defined categories. One cannot, for example, unambiguously determine the number of spills caused by ships pumping their bilges, because the encoder is given two choices, "pumping bilges" and "vessel underway," that might be applied. These criticisms are not meant to imply that the PIRS data is of little use, but rather are a warning that one must be fairly skeptical of tabulations or analyses in which these problems are not addressed.

The Geological Survey's Event file is also in coded form, although the record format provides space for a written narrative of the incident. The spill source identification is provided by area, block number, and structure number.

The structure number allows cross-referencing to the structures file, which contains details of the structure's age, size, and function. The pipeline incident file maintained at Metairie is along the lines of a simple manual tabulation. No codes are used. In working with the spill files we found that some of the incidents listed in the

Metairie file were not included in the Event file. The incident numbers in the Event file are not consecutive, and we speculate that some of the incidents were deleted due to revisions in the estimate of the amount spilled, since this file retains only those incidents of one BBL and larger.

The LPR10 production file lists the annual production of oil, gas, and condensate for the various leases. It is not structure-specific, and there is some problem in establishing which structures are aggregated within the lease categories. We used the LPR10 data primarily in conjunction with the Coast Guard's PIRS data for the purpose of establishing gross relationships between lease production and total spillage. Our difficulties with this database were documented in a previous report submitted in March of this year.

The Martingale Master Vessel File (MVF) lists the tankers and bulk carriers of the world based on a number of independent sources (see Appendix A). For our purposes we required the vessel call sign, age, deadweight and gross registered tonnage, name, flag, and draft. We also required the official registration numbers of U.S. tankers, which we obtained from the American Bureau of Shipping's Register.

Throughput data was acquired from the U.S. Army Corps of Engineers' Waterborne Commerce of the United States and from United Nations and U.S. Census Bureau tabulations. We were unsuccessful in obtaining detailed, flag-specific port call and journey data for tankers in U.S. waters. Our inquiries at the Census Bureau and the Maritime Administration did uncover the basis on which such information might be calculated. This is the Census Bureau's AE350 and AE750 tapes and confidential Corps of Engineers data. We considered analyzing this data ourselves, but the Census Bureau quoted us a delivery date for the AE350 and AE750 tapes in July or August, too late for our purposes. The Maritime Administration indicated that they were working on this problem and that the information required would be available sometime in 1978 for 1975 onward.

Appendix A provides a more detailed discussion of the data management problem.