TITLE: Federal Mineral Lease Bidding Data Bases Versus the Real World and Vice Versa

AUTHOR(S): John Lohrenz
             Ray A. Waller

SUBMITTED TO: Conference on Auctions and Competitive Bidding, December 6-8, 1979, Seven Springs Conference Center, Mt. Kisco, NY
FEDERAL MINERAL LEASE BIDDING DATA BASES VERSUS THE REAL WORLD AND VICE VERSA

John Lohrenz  U.S. Geological Survey
Conservation Division - MS 608
U.S. Geological Survey
P.O. Box 25046 - Denver Federal Center
Denver, CO 80225
(Tel. no. 303 234-5453)

Ray A. Waller Los Alamos Scientific Laboratory
Los Alamos, New Mexico
Group S-1 - MS 600
Los Alamos Scientific Laboratory
P.O. Box 1663
Los Alamos, New Mexico 87545
(Tel. no. 505 667-3308)

ABSTRACT

From the vantage point of builders and operators of Federal mineral lease bidding data bases, we address users. Users are a polyglot lot of empiricists, mechanists, and theorists. The users' "perfect" data base would contain all data needed to examine all bidding theories and predict bidders' behavior using the theories. We have not nor are we likely soon to attain that ideal data base. However, even the richly flawed history of these data bases leads one to the not particularly prescient notion that users, with omnipresent complaints about errors in and lack of data, will keep the data bases functioning by using them.

-------------------

This draft manuscript was prepared as a basis for presentation at the Conference on Auctions and Competitive Bidding: Uses, Theory, and Development, at Seven Springs Conference Center, Mt. Kisco, New York, December 6-8, 1979, under the auspices of the Project Center for Competitive and Conflict Systems Research at Cowles Foundation for Research in Economics at Yale University. Comments and criticisms about this draft manuscript are solicited; however, no proscription against citation or referencing of the manuscript in the current form is implied.
I. INTRODUCTION

Starting in 1971, we have been building and maintaining data bases about Federal mineral leases that are offered for competitive bidding. The data bases are a full census of the leases under consideration. By that, we mean the data bases include each and every lease in question and not a sample of leases taken either randomly or of those for which the data are available. If a lease occurred in the real world, we want to be sure that lease is in the right data base. Table 1 summarizes the status of the data bases. Note that the data bases include an array of minerals.

By far the most common method extant for offering Federal leasing has been sealed bonus bidding wherein the bidder offering the highest "front end" cash bonus may be awarded the lease subject to a pre-defined schedule of royalty assessed to any subsequent production. The data for bonus bidding in the data base includes the amount of the highest and any other bids along with the ownership of all bids bid. More recently, some leases have been offered under a sealed royalty bidding system wherein the bidder offering the highest royalty may be awarded the lease subject to a preset bonus. In this case, all royalty bids and their ownership are included in the data bases. Actually, the Outer Continental Shelf Lands Act Amendment of 1979 (16) provides for offering leases using a profit sharing or working interest bidding procedure as well as the bonus and royalty bidding alternatives already used. No leases have been offered by profit sharing or working interest bidding so far, however.
An intent is to keep the data bases current with annual updates. Table 1 shows that intent is only partially realized. Source data delays annual updates 2-3 months after the end of the year. The larger delays are due, simply, to the inertia in the process of getting the new data in the data base in usable form. Another intent is to include production data for all leases which had any production. Production data are included in the data bases only for Federal oil and gas and coal leases. There has been no production, so far, from Federal geothermal leases. Production data for Federal and Indian onshore oil and gas leases are not yet in the data bases.

All of the data bases of Table 1 include only publicly available data. The data bases are available to the public. (Contact either of the authors if you would like any of the data bases. We will send a magnetic tape with documentation about the data bases requested. This magnetic tape may be duplicated and returned.)

As may be true with many data bases, there was no stated intent to build these data bases as such when the work started. Instead, there were applications which led to the creation and maintenance of these Federal mineral lease bidding data bases. The first application was a request from management to determine the present values of the government revenues from Federal offshore oil and gas leases by type, bonus, rent, and royalty. The second application was a study suggested by a policy-maker considering a hypothesis that higher bonus bidders were "stockpiling leases for speculation"
and not for actual exploration. (Some of the technical results of this study and its conclusions are given by Lohrenz and Oden(14). The decision to build a suite of Federal mineral lease bidding data bases was really never made explicitly, but implicitly as these two applications and, then subsequent ones arose and were dealt with. One might presume that any explicit decision to create or not to create the data bases in 1971, knowing that, in 1979, the costs would have accumulated to near $10 million, would likely have been negative.

When we first started building these data bases in 1971, one organizational manager demurred from allocating any of his people to getting the needed source data out of files and archives writing, "We have never had any use for these data in the past and we will not in the future." Now, in 1979, we have transmitted over 400 of these data bases to requesting users. (In addition, we have transmitted many more analyses that we and others have performed using the data bases.) The users have included other Government agencies, contractors to Government agencies, public interest organizations, bidding companies, and consultants and companies providing services to bidding companies. We think it proper and of interest to note that the data bases have occasionally been sent to both parties of a legal dispute. Sometimes, one party has been the Government itself. These data bases, like all good data bases, are truly unbiased, but that does not imply that users are or should be. Regardless of users' bias, if any, these data bases exist and are justified on the basis of users they have. And because we have these
users, we can now confidently refute the organizational manager who predicted no users. He was wrong about that, but he was right when he recognized that users using data bases justify them. The fundamental reason we started building these data bases in 1971 is that we presumed users would use them and the reason we maintain these data bases on 1979 is that users are using them.

It is the community of users--of which we are a part--that we are addressing in this paper. While the community is large, it is disparate in character. Having seen so many users, we think we have some observations about some users that should interest other users. In Section II, THE THREE KINDS OF USERS, we classify users as empiricists, mechanists, and theorists and show their differences. The bidding data base empiricists, mechanists, and theorists would like to have is developed from considering how bidders might bid in Section III, THE "PERFECT" BIDDING DATA BASE. Where previously we have addressed the community of users as users, in Section IV, THE LAWS OF DATA BASES APPLIED TO MINERAL LEASE BIDDING DATA, we address users as operators of the data bases. Finally, we state an unsurprising prediction about the future of these Federal mineral lease data bases in the last Section V, THE END IS NOT NEAR.

II. THE THREE KINDS OF USERS

We classify users of these Federal mineral lease bidding data into three kinds: empiricists, mechanists, and theorists. Any general classifications like this oversimplify, but we think it useful for describing users
overall. Obviously, we are personalizing the three kinds of users; it is the users' applications that are of three kinds and some users have applications of more than one kind.

Empiricists use the data base simply as a source of data for observing correlations and lack of same between variables. This is done without postulating mechanisms of proposing theories. For empiricists, it is of interest and enough to know, from the data, that correlations do or do not appear to exist and, given a correlation, something of their form. The scattergram is a basic tool. Newer tools are useful for exploratory correlation searches on multidimensional data.

Empiricists are extant in other fields than bidding data. Geologists who map rock-types are empiricists. Empiricists, whatever technical field they are working in, adopt the attitude of the Tin Pan Alley songwriter who was satisfied with the correlation, "I don't know why I love you like I do, I just do".

Empiricists have one operational advantage compared to the other kinds of users. Given any data, they are never devoid of data to practice their craft. Regardless of what data are in the data base, empiricists have the requisite meat (or gristle) to look for correlations.

Many mechanists and theorists might note that the empiricists' operational advantage is that they can always look for correlations, however meaningless. In that way, mechanists and theorists impinge on
the disadvantage of the empiricists' product. The disadvantage is that any correlation, however well grounded in theoretical statistics, does not imply any mechanism per se between the variables correlated. Using any correlation for prediction implies some sort of assumption that all other things remain equal when one may not, in fact, even have any idea about what those other things are.

Defense of the meaning of any empirical correlation against attack is difficult; potential attackers have so many exposed flanks to bore in on that cannot be defended. For example, Lohrenz and Oden(14) presented histograms with a correlation showing that leases with higher bonuses per acre tended to get higher royalty per acre from production. Not so, said one declaimer. The lower bonus leases were leases the bidders thought were gas leases which had, the declaimer avowed, a product of lower value than the higher bonus leases the bidders thought were oil leases. Aha, you might think, let us check that by dividing the population into leases the bidders thought were gas and oil leases. But, those data were and are unavailable. So, our declaimer could not be answered from the data. One cannot defend a correlation against a mechanism imputed to be operational in the correlation that must be tested by data that are not available. That is one example of an indefensible flank where all empiricists soon learn to expect pummels.

Mechanists explicitly or implicitly start by presuming some mechanism. Providing the data required to test the mechanism are in a data base,
mechanists use data bases to "discover" the data fit the mechanism and, therefore, the presumed mechanism is in effect, they aver. Of course, mechanists are alive and well in all technical fields. A favorite example of mechanists-at-large in mineral lease bidding is the work assuming bonus bids on leases are drawn from a random lognormal distribution. As Crawford(5) pointed out, the mechanism purported to occur if bids are lognormally distributed is the bids are computed in a multiplicative process in a manner that, in concert with the central limit theorem, causes bids to be lognormally distributed. And that presumed mechanism is the springboard for mechanists "measuring" the lognormal distribution.

But, as theorists are quick to point out, the fact that the data fit, at least to some extent, what a mechanism predicts is not conclusive evidence that the presumed mechanism is, in fact, in force.

One is frequently able to ascertain a fundamental, possibly necessary, antipathy from theorists to the work of mechanists assuming the lognormal distribution applies and studying the properties of that distribution in the data(6,11,12). These theorists hold most ardently that it is anathema to common sense to hold that bidders arrive at any bid by some kind of random sample from a lognormal or, for that matter, any other kind of statistical distribution. Further, these theorists hold that, even if one could show the bids appear to follow some sort of random statistical distribution, that is a happenstance of no or, at best, passing interest.
What is of interest, theorists hold, is to be able to articulate a theory which shows how and why individual bidders bid the way they did from the data. Further, that theory should be sufficiently robust that one could confidently predict how individual bidders would bid under any and all alternatives for offering the leases or whatever for competitive bidding.

We repeatedly see and hear arguments between empiricists, mechanists, and theorists using (or trying to use) the Federal mineral lease bidding data bases we have. Generally, the arguments are productive although, sometimes, we see a certain chauvinism within an empiricist, mechanist, or theorist holding the notion that only his or her kind of application is appropriate. It goes without saying that the study of competitive bidding, like any other technical area, requires empiricists, mechanists, and theorists among ... no one kind is more elite than the other (or less prone to commit foolishness).

In these arguments between empiricists, mechanists, and theorists, one can see something akin to the Heisenberg uncertainty principle of physics operating, i.e., the closer one gets to the exact measure sought for, the more difficult the actual measurement becomes. Thus, the theorists' stated goal to be able to correlate past bids from theory and predict future bids under different alternative bidding situations is a goal theorists and every one else would like to reach, but, given the data at hand in any real world,
the goal is unattainable. So, the mechanists and empiricists work away toward lesser goals under criticism, sometimes unfairly, that they are not working toward the theorists' goal that is unattainable from the data bases at hand.

How could we make the theorists' goal attainable? What kind of bidding data base would we have to have (and can we get it)? These questions introduce the next section.

III. THE "PERFECT" BIDDING DATA BASE

The "perfect" data base of mineral lease bidding data that would allow the present and future bidding theories to be examined would have to capture the salient data that occur and arise leading to the bids that are bid. What are the data we would need to have a full set needed for this sought for "perfect" data base?

To answer that question, we want to contrive a peculiar kind of a poker. Poker is, in this culture, a symbol of gambling and anyone, even a government, offering to sell anything by competitive bidding is inviting (and hoping to benefit from) participants to gamble.

Let us imagine one table of poker players. We're playing five-card draw. Let's consider two hands dealt to you. The first hand dealt to you, before any drawing, is a pair of deuces, and the other hand, three jacks.
Actually to make this poker game like bidding for mineral leases with admittedly huge uncertainties involved, we should "peculiarize" this poker game by saying that you have only the word of a flawed estimator that you have a pair of deuces in one hand and three jacks in the other. (See Stark(19) wherein the importance of estimating value of the item offered for bid in the bidding process is particularly noted.) You may know something about historical biases and variances between actual results and estimates your estimator has made, but that is all.

Any poker player knows, of course, that three jacks, if you really have them, is a better hand than two deuces, if you have them. But, the three jacks aren't competing against the two deuces; each must compete against a different set of hands. Nonetheless, the three jacks is a more valuable hand in the sense that every poker player would prefer to have three jacks over two deuces ceteris paribus. Let us presume you are as prudent a poker player as mineral lease bidders are prudent. How would you bid the three jacks? The two deuces?

In poker, just like mineral lease bidding, you are aware of the potential range of outcomes. The draw might change the entire character of your hand and make it more powerful. You can win big with a relatively weak hand and lose big with a relatively strong hand. You know the probabilities don't favor that outcome, but the probabilities are finite. Obviously, the answer to how you would bid your three jacks or two deuces is, "Well, that depends."
On what? We have already identified one thing which will affect how aggressively or conservatively you bid your three jacks or two deuces. First, your bid will be affected by what you perceive to be the hand you hold and the possibilities of what that hand will become.

Second, since you are a prudent player, your bid will also be affected by your perception of what your competing players think they hold. That's why you may bet a weaker hand like two deuces more aggressively sometimes than a stronger hand like three jacks on the next hand.

Third, while you are a prudent player who would like to win as much as possible, you also do not want to risk having to leave the game. After all, you didn't get into the game to leave it. So, regardless of what kind of hand you have, your bid is affected by how much worry the bid gives you that you may have to leave the game. It's a poker player's goal, first, to feel safe and, then, to make some money playing the game. It's a bidder's goal also. In fact, one can consider it a societal goal. (Philip Slater(18) opens his book, "The Pursuit of Loneliness" with the statement, "One of the first goals of a society is to make its inhabitants feel safe.")

A bidder once confided to us a story of the consternation of his organization's board of directors after discovering they had spent a distressingly large portion of the current liquidity in one Federal offshore oil and gas sale. This followed the board's decision to be aggressive in
that sale. That would certainly be analogous to a poker player with a "bot" hand wanting to bid high, but worrying about the dwindling chips at his or her disposal.

Fourth, as you are a prudent player you have a memory and your bid is affected by your memory, by history. Because you have this memory, you would conceivably play two poker situations that have exactly the same "optimum" bid based on everything else you know differently.

Fifth, while you are a prudent player, that does not mean all other players agree that what you consider prudent play is, indeed, prudent. Perhaps, you might enjoy playing poker for the camaraderie of the players. At least, you may enjoy it as long as you don't lose too much. Perhaps, you want to exhibit a certain flamboyancy and bid aggressively. Or, perhaps, you want to exhibit a certain flamboyancy and bid aggressively. Or, perhaps, you want to exhibit some sort of somber mien and bid conservatively. It is doubtful if there is any poker player who, at least a few times, did not bid in order to display something that was not there, be it a hand or wit. The point is that poker players, like mineral lease bidders, can have different measures of what counts as success in their gambling.

In summary, we have identified five considerations affecting what a poker player will bid on a hand and a mineral lease bidder for a lease. These are:
For the poker player

1. The hand you think is being held and the expectation of what that hand may become.
2. The hands you think the competing players are holding and what they might become.
3. How large or small the stack of chips you have before you is.
4. Things that have happened before in this kind of game with these players.
5. The things you want out of the game.

For the mineral lease bidder

1. The estimate of what the lease has and what value might come from it.
2. What you think competing bidders think about the lease and what they might do.
4. The previous history of what happened bidding on leases.
5. The things you hope to get done by bidding.

All of these considerations would affect how you bid in a poker game and for mineral leases being offered.

To make the poker game analogy to mineral lease bidding analogy even more complete, we should really consider a mineral lease bidder as a participant in a number of tables of poker simultaneously. The multi-table poker table may, in fact, seek to be entering additional tables of poker or, for that matter, seek to leave some tables. That such intentions may affect the poker player's bids on one table is obvious. A mineral lease bidder once attributed his organization's aggressive bidding in a certain sale as caused by the expectation of certain expropriation payments. That would be like a poker player bidding higher on one table because he expects to be told to take his money and go elsewhere on another.
Unlike poker where a player on many tables could transfer winnings from one table to another and, at any time, compute a single number of the value of his holdings on all tables, the game of mineral lease bidding does not allow such simple machinations. That is one way in which the poker game analogy to mineral lease bidding breaks down.

Mineral lease bidders are like poker players playing at simultaneous tables that may have strong physical relationships between tables, i.e., there may be a physical benefit to be playing at a specific set of tables such that a loss threatened on one table calls into question the safety of more than just the table in question. A mineral lease bidder who has favorable contracts to deliver finished products to a certain region that are threatened by inadequate supplies may be more aggressive in a bidding game for leases that could alleviate the threat.

If these considerations that would affect poker players' bids also affect mineral lease bidders' bids, it is not surprising that bids may appear to have arisen out of some random distribution even though no bid was random. (This appears also to be why it is difficult to ascertain one statistical distribution which clearly describes the bids, e.g., see Bruckner and Johnson\(^7\) who studied the lognormal, log-uniform, and Beta distribution.)

Let us return now to the question at the start of this Section. What are the data we would need to have in a "perfect" data base to seek the
bidding theorists' goal of correlating bidders' bids in such a way that we can predict future bidders' bids? Consistent with the five considerations of our poker player and mineral lease bidder, we would need:

--each bidder's estimate of the value of the lease in question
--each bidder's estimate of competitor's thoughts and tactics
--each bidder's liquidity that would be affected by the bidding
--each bidder's recorded history of successes and failures on leases of the type in question
--each bidder's stated goal or function which is being optimized

In addition, we would need, for each bidder, the status of all other business enterprises the bidders are involved in, wants to be involved in, is seeking to get out of, or is being forced out of. Further, all of these data would have to be available in some fully structured quantitative format.

We do not, of course, have these data in our current inventory of Federal mineral lease data bases. We do not want to jump to a conclusion that it would be completely impossible to get such a "perfect" data base, but it certainly would be difficult anytime and is impossible under current policies and procedures.

What bidding theorists along with empiricists and mechanists should renew, we think, from noting the gap between the "perfect" and actual data bases is a sense of humility. The Federal mineral lease bidding data bases contain, at best, only a modicum of the relevant data that exists in the
real world. Theorists, mechanists, and empiricists are all in the "bush leagues" in the sport of correlating and predicting bidders' bids. There appears to be little reason to bicker about who is more and less "bush".

IV. THE LAWS OF DATA BASES APPLIED TO MINERAL LEASE BIDDING DATA

At the outset, we noted that the audience of this paper was the community of users of Federal mineral lease bidding data bases. In Sections II and III, we have been addressing that audience, but as users change perspective in this Section and address users as operators of data bases,

We will first state three laws applicable to all data bases—not just mineral lease bidding data bases. These three basic laws about data bases are exactly analogous to and, in fact, drawn from the three great laws of thermodynamics. Like the laws of thermodynamics, these data base laws state constraints, limits, and unavoidable burdens. Also like the laws of thermodynamics, there is no positive proof that these laws are correct excepting that no one has been able to violate them. Here are the three laws of data bases:

First Law:

One cannot get information based on data that is not at access in a data base.

Second Law:

When there is questionable or apparently erroneous data in a data base, the only way the apparent problem can be resolved
and, if necessary, corrected is by doing work outside the data base and using that work on the data base.

Third Law:

The only data base which does not have questionable or apparently erroneous data is a data base that is not being used.

We will leave it to the reader to reconstruct, if necessary, the three laws of thermodynamics which these three data base laws emulate. Both sets of laws have, after their statement, an aura of simple obviousness that seductively mask an always abortive quest to commit exactly what the laws prohibit. Using these three data base laws which we will identify as the first, second, and third laws, we want to tell some tales about users (and people who wanted to be users) of the Federal mineral lease bidding data bases. The stories follow:

Tale No. 1: Data Doesn't Have a Family Name

One potential user asked us, "Why don't you have any environmental data in your data base?" As it happens, we did have a user who used the Federal offshore oil and gas data base production data to look for correlations between the amount of production and spills as part of a study on that environmental problem. Nonetheless, another potential user opined that the data in the data base "were not responsive to environmental concerns." These potential users, of course, failed in a fundamental understanding of what data in data bases are. Data are not classed by hue and color of any one person's particular application. There is no such thing as environmental
data per se just as there is no such thing as anti-environmental data per se. The first law notes there is only data without qualification which one can use to mold some kind of information which may be environmental or anti-environmental—which is generally a non-technical value judgment. Data cannot be politicized; information can.

We are always looking for comprehensive sets of data that we should add to the Federal mineral lease bidding data bases. If we are prudent data base operators, we should be. One must be guided, however, in adding data by the value of the value to users compared to the cost of adding the data and not the "brand name" of the information the added data would supposedly enable.

No data is categorized by any particular application it might be used for. Data doesn't belong to any particular application, family of applications, or family— from which we get the title to this tale.

Tale No. 2: Data That's Not

We are frequently asked why we do not have or get reserve data in our oil and gas data bases. Reserves, in that context, means the reserves which, in the future compared to some datum date, will be recovered. The answer is distressingly simple. We do not have future recoverable reserves in the data bases because future recoverable reserves are not data. They are all estimates about which there is no agreement except that all
estimators agree they are flawed and subject to huge uncertainties.

One can use an algorithm to estimate future recoverable reserves. The simplest example would be the extrapolation of established production (13). But even this simple procedure does not yield data but an estimate for recoverable reserves which should not be used without considering how it was obtained.

Perhaps it would be of interest to get data in the database of Estimator A's future recoverable reserve estimates all done in a specified procedure and at a specified time, actual or relative to leasing. Then, one could do a study of Estimator A's estimates of future recoverable reserves against what actually did happen. Or, if we also had data in a database of Estimator B's estimates with Estimator A's, we could compare A and B against each other. We must admit that such a comparison of estimator's estimates may not be the most meaningful or interesting application. (One estimator we know has called such comparison "a race to see who is slowest").

The most meaningful and interesting studies using the database treat the hard data most directly. By hard data, we mean data that is either right or wrong and not subject to how good or how bad we find out the estimate or the estimator who made it was. The bonus a bidder offered is hard data. Either the number is right in the database or it is wrong. That is not true about estimates of future recoverable reserves.
Users need to understand the differences between values that are data and that are not such as estimates and values connected, however tenuously, to the data. Sometimes one is tempted to state a corollary to the first law noting that one cannot get information based on data that is not data.

**Tale No. 3: Seek What You Will Define**

The single type of application that has attracted the most use and users of the Federal mineral lease bidding data bases have dealt with the broad question of competition. As one user put it, "We want to use the data base to determine if there has been competition or not." That is a clear and simple intent. It is certainly a laudable intent. But, is it a technically complete and realistic intent? We think not. Perhaps this is why that, when it is necessary to be tactful with users, we have memorized the stock riposte to that intent, "Please recognize that it is unlikely that you will be able to use the data bases to get a definitive answer to what you seek." When we can, we ask the user what is meant by competition. Sometimes the user responds with a definition that can be directly generated from the data in the data base. Concentration ratios are a specific example(2). We point out and some users admit, however, that any single measure of competition, be it concentration ratios or something else, is an incomplete and flawed measure. But, at least, these can be computed from the data and, hopefully, will be used properly with their burdens to meaningful interpretation realized.
Many users, however, somehow remain incredulous. Just like the medieval "thinker" who thought the laws of thermodynamics making perpetual motion machines impossible were some kind of conspiracy to hide their availability, so the assessment that the question of competition is unanswerable from the data is viewed by some as a kind of obstruction erected to hide the answer.

We have tried—only occasionally with success—to surreptitiously instruct these users in the first law by asking, "What would you expect the data to do if there were no competition?" Or, the converse, "If there were competition?" When those questions work, the realization dawns that there is no such thing as no competition and no such thing as 100 per cent competition. Anyone's definition of competition implies an amount, deemed adequate, of desirable competitive activities going on. What the desirable activities are and the amount deemed adequate are policy value judgments about which technical assessments cannot be made unless the activities are quantitatively defined. When the activities are quantitatively defined in such a way that the data allows their evaluation, the data base can be of assistance. The apolitical data base cannot, however, itself serve in any promotion or defense of the policy inherent in any specific definition of what constitutes desirable kinds and amounts of competition.

Tale No. 4: The Action Is in the Kitchen, but Few Want to Go in There

Users see the data bases as a source, hopefully, of answers to their application questions. That is natural. What users substantially fail to
The expense of time, money, and effort that must be expended to build and maintain a data base. Building and maintaining the Federal mineral lease bidding data bases has cost, to date, almost $10 million. By any standards, these are not huge data bases, but the costs of getting the source data, putting it in the data base, and checking the data base to try to find and correct the more egregious errors, a continuous process, are high. We should point out that the $10 million approximation does not include reports, many times of use in themselves, which were generated in the process of checking the data bases.

Users are not the only ones who sometimes exhibit a proclivity to ignore the process of what gets them the data they use in the data base. Systems designers have been guilty as well. An early effort to computerize these data bases involved a complete system implementation without ever having looked at the data per se that would be in the data base! Oh, there were many meetings with potential users who were asked what their uses would be. The system designers were disappointed that the users could not give specific uses. They hadn't seen the data either! But, the system designers spent much time and money selecting the hardware and software systems which would optimally process the data they hadn't seen for users who hadn't been able to tell them what they wanted! When the systems designers sent the people to encapture the data for their data base, the predictable wail arose, "Your data doesn't fit our system." There was a considerable fuss. You guessed it. The data won and the system lost.
Whether you are a system designer or a user, you cannot avoid getting your nose into the process of how the data got there. Some might seek the goal, as the system designers of the previous paragraph did, of designing the data base so its maintenance is transparent to users. But, the goal is a delusion. You cannot divorce the proper use of a data base from knowing how the data got in the data base just as an epicure cannot judge the content and quality of a dish purely by its presentation on the dining table. The process in the kitchen where the atmosphere is somewhat more hectic than the dining room has greater direct effect on the resulting dish than the flourish with which it is served.

Indeed, the corollary to Murphy's law applied to building and maintaining of data bases is that the source data for your data base is always in worse shape than you thought it was. Source data for a data base has a habit of being discovered as non-existent when one finally wants it in hand. The second law prevails. Data does not come from the real world to the data base and spontaneously find an appropriate notch to repose in. And the second law eats up a lot of time and money.

Table No. 4: Thank Goodness for Little Complaints

We have received a lot of complaints about apparent errors in the data base. We still get them. Because we use the data bases ourselves, we have done much complaining ourselves.
Some of the apparent errors turned out to be real and embarrassing. Early in the life of the data bases, we found that almost 100 percent of the royalty due to one Federal offshore oil and gas lease sale came from one lease of many producing. Reason? A shifted keypunch field which credited a lease with $10 million royalty in one year when the lease actually had none. (Another lease actually had $34,331 royalty that year, but showed up in the data base with $3433.) The data base system had all kinds of safeguards against that kind of error, but somehow the error got through. In this case, we could find the error because it led to an observable oddity and correct the error by checking through the system.

Most errors aren't that simple. Most errors have to be tracked down by questioning the source data you have and going beyond that to better source data. Sometimes the source data is odd enough to be funny. We found source data indicating a well had a true vertical depth of 68 feet and an actual depth of some 6,000 feet. Believing those figures to be beyond the state-of-the-art of even the most precision drillers, we checked and found the correct source data. The data base had to plug and abandon the "corkscrew" well. That is the kind of error that should not and did not, as far as we know, lead to any error affecting users. One can be exculpated for such errors.

But, not so for other errors such as the discovery of leases in the data bases with substantial production, but without any wells drilled in
them as well as leases in the data bases with no production and many wells drilled in them. The former is physically impossible; the latter is wholly unreasonable. Both proved to be errors corrected only after going beyond the original source data. This is an example of a kind of error in the data base which can be pernicious. A user studying the profitability ranges of individual leases can get results that, on the surface, look correct, but are grossly incorrect. The potential for technical foolishness hidden in results is enhanced in this case, as in many situations, by what may be called the "interest and error in the extremes principle". So many times, the user is interested in some highest or lowest extreme values of some computation based on a data base and errors in a data base seem to most affect highest or lowest extreme values of a computation from a data base.

Some of the apparent errors turned out not to be errors given the specifications for the data in the data base. An example that arises frequently is the "error" of negative royalties and production for some leases in some years. Physically, that appears most unlikely. Physically, it turns out, that did not happen, i.e., the government did not pay a leaseholder royalty nor did a leaseholder "inject" production into a lease. Federal and leaseholder accountants can make adjustments, however. Sometimes these adjustments are made at the behest of a court order or decision. The adjustment may involve an overall accounting settlement involving many leases; the adjustment to any one lease may be entirely arbitrary having nothing to do with any actual production and royalty from the lease. Such negative royalties and production may confound users who want to use the data bases to obtain
solely physical production quantities with time. The user must either proceed with the nonsense of negative physical production or invoke some kind of algorithm to avoid the nonsense. The object of the data bases is to capture the production and royalty credited to a lease. These are data which can be checked. As data base operators, we might wish that accountants would make adjustments otherwise. Unfortunately, accountants (and those who make juridical decisions) seem uninterested in easing the work of data base operators in their work and it does not seem one should plan this will change in the near future. We are, in this way, saddled with apparent errors that are really not errors, albeit we wish the source of the apparent errors wasn't there.

Sometimes the apparent errors turn out to be a misunderstanding of the meaning of the data in the data base. An example is well completions. The LPR-19 data base shows wells either as completed or not. Based upon the unadorned meaning of the word, "completed", the preemption might be that a completed well was a "successful" well, i.e., a well which was sufficiently productive to meet some "success" criteria. But, not so. The specifications are that completion is simply a kind of mechanical process done for a variety of reasons not necessarily leading to (or seeking) production. Precisely what constitutes a completion has no precise definition or standard. A completion is what the individual lease operator reports as a completion. No source data exist about the purpose and intent of an operator's completion or its ultimate outcome in any structured format or specification. This is
an example of an apparent error which is, in fact, not an error, but a constraint in the data bases imposed by the source data. The "error" cannot be corrected without source data not now available taken according to standards not now invoked are at hand.

We want to mention one array of errors discovered in the LPR-19 data base involving completions to illustrate one of the more troublesome facets of a data operator's plight. The LPR-19 data base included individual well data without completion data originally. We decided to add completion data subsequently. After we added completion data, we found something peculiar. The data showed that, overall, completed wells clearly tended to be the earlier wells drilled on a lease. While that could happen sometimes, that was not an expectation given the usual scenario of exploration drilling. You can imagine the consternation when we discovered that when adding the completion data, the addition had been made without regard to which particular well in a lease had been completed. If a lease with 21 wells had 17 completions, then the first 17 wells were taken as completed of the 21 in drilling date sequence. Those who handled the data said quite righteously, "You did not tell us completion data was for individual wells." The data base operators replied with equal righteousness, "We didn't think it necessary to specify that completions apply to particular wells." After wasting a little time and energy in that game of mutual reproach, we started over with adding the completion data. The second law applied again.
Especially for competition studies, one would like to be able to consistently identify bidding organizations. Just like a bettor at a race track examining the past record of a particular horse, one would like to examine the past record of a particular bidder. Users try to do this and we try to help them with codes in the data base, but the results are never wholly satisfactory. There are basically two reasons. The first is mergers, purchases--whole and partial, and other corporate arrangements which leave the decisions as to who was really a distinct bidding organization separate from all others murky. The second is joint bidding. Dougherty et al(10) show some of the joint bidding situations which cloud the definition of a distinct, individual bidder. Now, what users would like to have is not just a current list of distinct, individual bidders, but one for any arbitrary time. (We are seeking to add assignment to the data bases in order to facilitate this.) However, it is clear that it will always be ambivalences and arbitrariness in any definition of distinct, individual bidders.

Some users see "mistakes" in how the data in the data base is arranged. For example, when more than one owner participates in a bid in a joint bid, the tape and card formats of the data bases include the bid value only once for all owners. For some users and their applications, it would be more efficient to repeat the bid value for each owner. But for other users, that would be a redundancy and require reprogramming of their applications. As data base operators, we see the plethora of user applications and know that
no single data base format can be optimum for all or even most applications. Conceivably, one could do a census of applications and arrive at some optimum format for the aggregate, but we think that hardly justified given the constantly changing mix of applications and our inability to predict the future applications.

As every data base operator knows, the cycle that keeps a data base alive is new applications leading to new specifications for the data in the data base leading to more new applications and so on. The best, the necessary evidence that the cycle is, indeed, going on is users' complaints. That is what the third law states. If we didn't have users complaining, we would be dealing with a "dead" data base.

V. THE END IS NOT NEAR

Economics has been called the dismal science. The three laws of thermodynamics and their analog applied to data bases portend, no doubt, an even more dismal viewpoint. Students, quite accurately, sometimes remember the three laws as: (1) You can't win. (2) You can't break even. (3) You can't get out of the game.

In Sections II and III, we have shown that, at best, any data base is that subset of the real world needed to find the solution to your riddles. And we saw that we have not captured the requisite data to do that in the Federal mineral lease bidding data bases. In the games between our data bases and the real world on the real world "court", we've lost, so far.
Immediate prospects for winning there aren't good either.

In Section IV, we recounted some of the games we've played against the real world on our data bases' court. The real world has been winning there, too.

But unlike the won-loss record of competing teams, you don't count data base wins and losses, but only that you could and did keep playing the games.

It is, after all, the purpose of these data bases to obsolete themselves. One can imagine investigators maintaining some sort of data base recording the times of sunrise and sunset long ago. Then someone came up with a model which made maintaining that data base obsolete. (No doubt, somewhere even today physicists are maintaining such a data base with time measured with modern day precision and "revising the model"). That made a data base of sunrises and sunsets recorded by hours and minutes obsolete.

So we, too, consider the prospect of having obsolete data bases of Federal mineral lease bidding data on our hands. But, the end does not appear to be near and, until that dismal end, we shall, no doubt, lose many more bouts with the real world.

ACKNOWLEDGMENTS

Russell G. Wayland and Hillary A. Oden of the U. S. Geological Survey were proponents of these data bases at their inception. The initial
applications were principally from Robert H. Lawton, then with the Department of Interior and now with the Department of Energy, and Hubert M. Risser, then with the Survey now deceased.

Kenneth A. Douglas, Robert S. Martin, and, later, Frankie W. Conner were principals with the General Services Administration in Fort Worth, Texas, where most of the seminal work with the Federal offshore oil and gas lease data bases was done. R. Gordon Peterson, then of the GSA Denver office, and Joan L. Keiper of PRI Services were mainly responsible for creating the onshore analog of the offshore oil and gas lease data base. The building of the LPR-19 data base was done mostly with Michael H. Seeb, then with CSC Infonet. Joyce Core of the Survey built the coal data base starting with finding the source data.

The LPR-17.3 data base for computerized maps of offshore leases was done with John L. Sibert of the Los Alamos Scientific Laboratory as a principal. Others at LASL who have had major roles in developing these data bases were Lawrence A. Bruckner and Myrle M. Johnson.

REFERENCES CITED


Table 1

Federal Mineral Lease Bidding Data Bases:

Summary of Status

<table>
<thead>
<tr>
<th>Data Base</th>
<th>Leases in Data Base</th>
<th>Time Period Covered</th>
<th>Includes Production Data</th>
<th>Documentation References</th>
</tr>
</thead>
<tbody>
<tr>
<td>LPR-5 &amp; -10</td>
<td>Receiving any bids</td>
<td>1954-1978</td>
<td>Yes</td>
<td>(3)</td>
</tr>
<tr>
<td>LPR-19*</td>
<td>Issued</td>
<td>1954-1976</td>
<td>Yes</td>
<td>(21)</td>
</tr>
<tr>
<td>LPR-17.3*</td>
<td>Receiving any bids</td>
<td>1954-1976</td>
<td>Yes</td>
<td>(17,22)</td>
</tr>
<tr>
<td>LPR-18</td>
<td>Offered</td>
<td>Mid-1972-1977</td>
<td>No</td>
<td>(15)</td>
</tr>
<tr>
<td>Indian oil and gas</td>
<td>Offered</td>
<td>1976-1977</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>State of Alaska oil and gas</td>
<td>Offered</td>
<td>1959-1974</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>Federal geothermal</td>
<td>Offered</td>
<td>1974-1977</td>
<td>None</td>
<td></td>
</tr>
<tr>
<td>LPR-20</td>
<td>Issued</td>
<td>1965-1978</td>
<td>Yes</td>
<td>(1)</td>
</tr>
</tbody>
</table>

*LPR-19 Includes all the data types included in LPR-5 and -10 plus individual well and platform data.