"DATA" The other four letter word

Terry Sinton, The Excalibur-Gemini Group

Summary
E-mail, E-Commerce, E-Business, E-Banking, the buzzwords of the 21st Century. What's Next? E-Data? E-Gad! The reality of the E-world we are plummeting towards is electronic data. Are we ready to go there? How do we get there from here? "Data" regardless of its industry of origin has accumulated over a period of time during which the "standards" for collecting and storing have changed. Varying media types, deteriorating physical condition, storage conditions, incomplete datasets are elements which effect data quality and determine on-line readiness. Preparing for electronic data is a complicated process that requires all the stakeholders to be active participants.

Introduction:
Let's begin by defining some terminology. Webster's dictionary defines "data" as "factual information used as a basis for reasoning, discussion, or calculation". Data in this paper refers generally to this information regardless of its form, structure, content, industry of origin, or end use. In other words data could be a seismic record section, a medical x-ray, a legal contract, an invoice, a well test, a land agreement, a geological map, an annual report, or any piece of information which has value enough to be kept for some period of time. Webster's defines a "database" as "a collection of data organized for rapid search and retrieval". For our use that definition works well. Webster's defines "on-line" as "being controlled by a computer" and for our uses "on-line" refers to data which is accessed only via computer. Finally archive is defined by Webster's as "to file or collect documents or items into a repository". For our purpose the definition of "archive" will be " to file or collect documents or items into an electronic repository".

Now that we have the terms of discussion defined let's look at the real question. Is "electronic data" really the future? While no one can say for certain, examining the trends of the last ten years gives us a clear indication of where we think things are headed. E-mail is a reality in most of our lives and has replaced many conventional modes of communication. The daytimer has given way to the palm pilot. Banks are discouraging customers from standing in teller lines by offering banking on-line. Investment companies are offering stock trading on-line. Organizations such as ARMA and the Junior Chamber International Senate are using on-line technology to not only contact their members but to communicate newsletters and meeting minutes. Storage of your personal photographs on-line is becoming commonplace. Music and literature are available on-line via the Internet. The internet if you believe the hype will give you information about anything you ever wanted to know and perhaps even a few things you wish you didn't.

There are many examples of business and government trends toward the transmission and storage of information electronically. In Calgary the Regional Health Authority has announced its intention to make all its records electronic, Oil and Gas companies are taking their seismic data online, the Daily Oil bulletin and local papers are available on-line, and the Alberta Energy and Utilities Board has just announced that all Well test data will be converted and made available on-line.

All of these examples along with the Internet craze give pretty clear indications that "data" is going "electronic". In our industry, the three most common driving issues that determine whether or not a company does to E-data are protection of deteriorating physical assets; the need to be more productive; and to decrease the amount of time required to access data. If it happens in your industry how does that effect you? Regardless of what industry you are in or what position you hold, you will be affected by E-data. How you will be affected depends on you.

Example: The impact of "E- Data" on the Geo-Scientist
"E-Data" is not new to the geo-scientist. For some time now certain types of data have been made available in electronic form. Stack data, and certain other types of data are commonly accessed through workstations. These examples were the beginning of the "electronic data trend and this trend preempts a future where all data will reside in an electronic warehouse.

Here are some of the questions the geo-scientist may want to ask about that future.

- What data will be archived? Some decisions are easy; for example no one will argue that field tapes are unnecessary. However there are other products which may be considered important to some individuals and not to others. For example fathometers, boat sections, drillers, shooters, chaining, sketches, tape logs, monitors, skid sheets, moveup logs, gather tapes, and brute stacks are all items which some geo-scientists may want kept while others do not.
- How many versions of processed products should be kept?
- Do you feel that all processed versions should be retained or only the last 2 or three generations. Maybe you want all processed versions kept only in tape format, or perhaps you feel that only the latest and greatest versions should be retained.
- Should interpreted data be archived?
- No one has a definitive answer to this question. Some companies want not only to keep some interpreted data, they want to keep **all** interpreted data. Others are considering keeping only the last versions of interpreted data, while still others will chose not to archive but to look at each area fresh. The answer may lie in the question "Is interpreted data a valuable knowledge asset?"

- What format and medium is going to be used? Is this data going to be readable by all of your service providers, and partners, and should you decide to change providers will you be able to transfer this data to the new system or will you be held for ransom by exorbitant translation costs?
- What data needs to be full on-line? near-line? off-line? This decision could drastically change the cost of your electronic warehouse and effect your exploration budget.
- Who is going to be in charge of the archiving project decisions? Are they qualified to make decisions, that will effect the quality of your data? Will they ensure that the geo-scientists are consulted before making decisions, which will have a direct effect on data assets? At the very least there should be a dedicated team whose sole responsibility it is to ensure the transition to E-data happens as it should. Do not be lead to the belief that cheap labor can be trained to manage this process. Your assets are too valuable to be entrusted to inexperienced workers who may not realize the legacy of their work.

E-data is here now. The cost of not participating in the process is that you may find your job is hindered by the actions taken and processes created. Your ability to meet land sale deadlines, locate oil and gas, and drill wells may be at risk. You understand better than anyone does what the day to day impact will be on your job so you are the best decision-maker.

Additional Concerns
In addition to the questions already discussed, your input is also of vital importance on some additional topics, which are critical to the success of an E-data project.

Seismic Sections
It is no longer necessary for geophysicists to use paper seismic sections to interpret data. The advent of workstation technology has made physical sections almost obsolete. We do still use them however, to quality inspect data for data sales and in data rooms for divestitures, and partnership deals. There are two schools of thought on how to reproduce seismic sections electronically.

One obvious answer is to scan sections into the system. The concerns with this solution are:
How many versions do you scan and maintain?
Do you replace old versions as reprocessing is done or do you just add?
How much storage space will be taken up by sections and at what cost?
How do you QC the scanned images?
Advantages: Scanning is relatively fast, and inexpensive

Disadvantages: Storage costly, maintenance can be cumbersome

The other commonly used option is to remaster your stack tape to complete the tape headers and add a file that contains label and header information. This method allows the user to send the stack data and file to the plotter and create a section at will. Some concerns with this approach are:
Do you remaster all stacks or just the latest and greatest version?
What do you do if you are missing a stack tape that matches a section?
Advantages: Sections are eliminated reducing processing costs, electronic storage is minimal, and multiple versions are handled in the management of the stack tape policy you adopt.

Disadvantages: Slows down the archive of historical data.

Basic Data
E-data requires scanning of basic data to tiff images. Historical data can be very difficult to scan depending on the media on which they are stored. Your basic data is probably on several different mediums including, paper, microfilm, microfiche, and floppies.

While scanning of paper images to tiff files can be more expensive than scanning microfilm or fiche the reasons to consider this are apparent. It is always best to scan an original document when it is available. Secondly scanning paper allows for the operator to adjust machine settings to accommodate variations in data quality. For example survey notes written in fine print with an “H” pencil require different settings than a chaining report written in black fine tip felt pen.

The most cost-effective process for scanning film and fiche is to use the first frame to determine machine settings and then run the entire role of film or sheet of fiche at that setting. The problem is that each frame may not scan clearly at that setting. The quality variances found in original documents, as well as quality control procedures used to create the film, can create drastic differences in image quality from frame to frame. This process requires stringent quality control of the scanning process and examination of each tiff image to ensure quality. Poor images need to be flagged and redone. This is not a problem unless you have millions of images, which some companies do.

One alternative to scanning microfilm and fiche is to scan all original paper first and for lines that have not paper originals, running of reader prints and using these for scanning can be a viable alternative. In this process each image is scanned separately allowing for adjustments in machine settings. While quality control is still required the amount of rework is usually significantly lower.

Go Forward Plan
While your archiving team is working on your historical data it is important to remember that each day your teams may be adding new data to your database with purchases or new shoots. It is vital to plan and develop the policies and procedures for the "go forward phase" for new data before the historical archive begins.

These policies and procedures need to include format standards for tapes, tape labels, transmittals, unique numbering systems, data entry, etc, which will ensure that new data fits into your archive and does not require rework. It is also important to establish efficient and effective workflows that protect the integrity of your database and the data itself. Choreographing this data ballet requires complete understanding of the data and who needs it when.

Conclusion
In Bill Gate's book "Business @ the Speed of Thought" he suggests that this century "will be about velocity: the speed of business and the speed of change". **E-GAD!** How many of you feel we are already there?

The goal of E-data according to Mr. Gates is to make your data "readily accessible".., "better organized".., "fresh and pertinent".. and "available to empowered workers". As stake holders you have the opportunity is to take control of the decision making process. Your gain will be **E-NORMOUS!**