ANATOMY OF AN EXPLORATION AND PRODUCTION DATABANK: PRACTICAL CONSTRUCTION TECHNIQUES

BY : UGUR ALGAN

(Published in the Leading Edge Magazine of SEG, July 1997)

This paper describes some practical approaches to designing, building and successfully deploying an exploration and production (E&P) databank.

Given the fact that the potential future assets of an oil company are only indirectly quantifiable through geophysical, geological and other engineering measurement methods, the market value and the competitive positioning of an oil company is a direct function of the quantity, quality, integrity and therefore the usability of its data assets. An E&P databank is a first step towards optimal utilisation of these data assets.

Historically, data management and related issues have been relegated to the domain of uninteresting issues which involve essentially a series of cumbersome cleanup, transformation, loading and reformatting steps. These steps were mere inconveniences in the days when our data assets were measured in terms of megabytes to gigabytes and the data domains were quite limited. However, they have now become serious stumbling blocks to progress since the size of data sets have grown to terabytes and petabytes, and we tend to show interest in many diverse data domains.

E&P databanks are now recognised as a significant element of effective data management. The construction of an E&P databank consists of the following distinct steps, each step providing added value in its own right as well contributing to the eventual outcome :

1. identification of the source, structure, usage and quality of all existing data assets;

2. definition and execution of a physical data upgrade programme to deal with physical deterioration and indexing issues;

3. identification of a target data model architecture, and definition of the model mapping from source data sets to the integrated target (termed logical data upgrade);

4. cleanup of data and transfer to the target;

5. definition of data exchange mechanisms with application suites;

6. definition of ongoing data acceptance policies, formats and procedures;

7. definition and implementation of the target hardware and communications environment as well as the system and data management environment.

Each of the above seven steps requires careful planning, well monitored execution, involvement from the appropriate stake holders, and the right skills mix.

Steps 1 and 2 have traditionally been regarded as physical data upgrade programmes which involve transcription of tapes, scanning of documents, and indexing and cataloguing of all physical data assets. Numerous projects of this nature have been undertaken throughout the world over the past five years. Business drivers for these projects have been preservation of data assets and improved access to data.

Steps 3 and 4 have received much focused interest, especially in the past three years with the emergence of Petroleum Open Software Corporation's (POSC) integrated E&P standard data model named Epicentre. The key drivers for these steps are better data utilisation through data reconciliation and integration, and better data management practices. Although projects of this nature have resulted in better data management environments and improved response time to end user requests, the direct involvement by the end users has been limited given the fact that most users are only interested in data to the extent to which they can be utilised by the end user's applications.

Step 5 involves a variety of manual, semi-automated or automated exchange procedures. Steps 6 and 7 are general housekeeping issues.

Each one of the above steps is an incremental advance towards the higher goal of optimal data management. Figure 1 shows the increasing value of the introduction of better data management practices, coupled with an effective IT infrastructure.

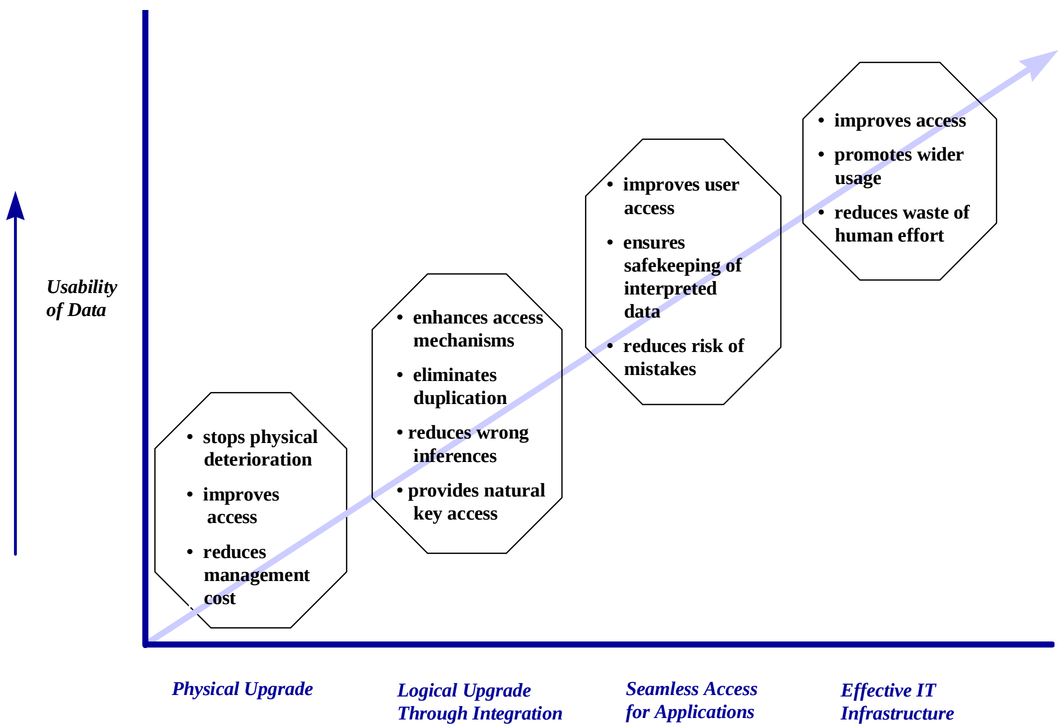

***FIGURE 1*: BETTER DATA MANAGEMENT PRACTICES LEAD TO MORE EFFECTIVE DATA USAGE.**

Figure 2 shows the complete flow of data from capture operations into the databank and out to the applications, annotated with the particular step numbers described above.

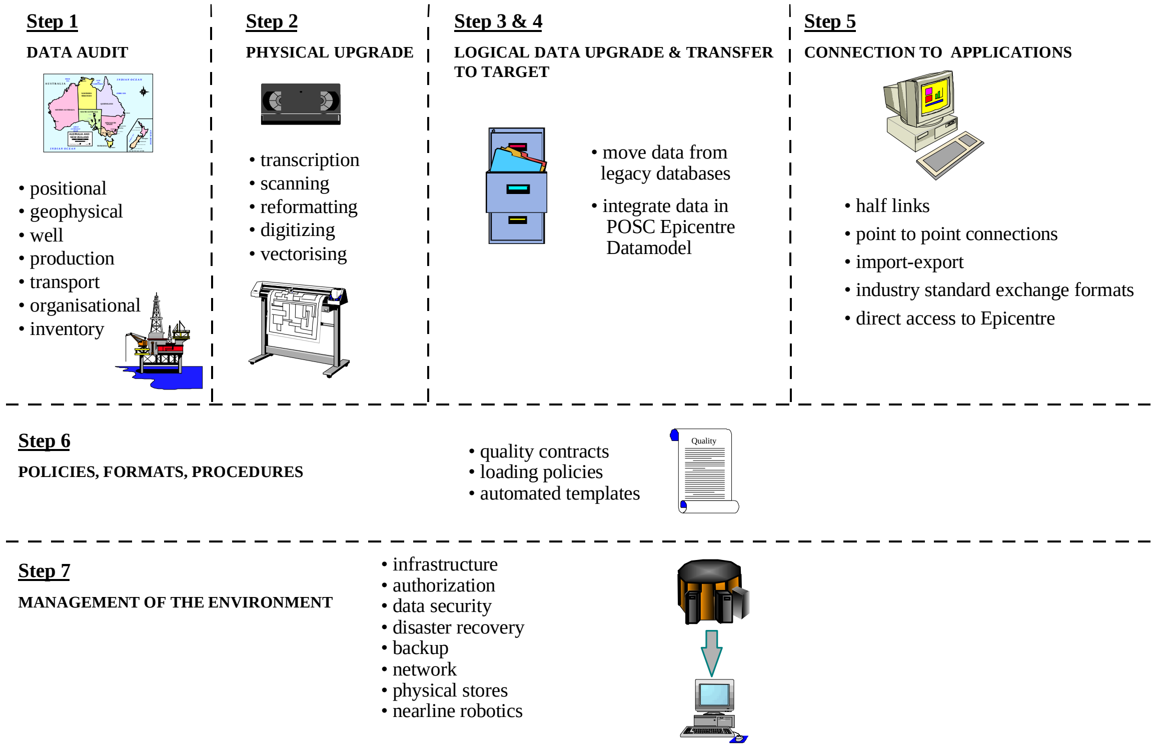

***FIGURE 2*: FLOW OF DATA THROUGH COMPONENTS OF A DATABANK**

The particular data domains involved in a databank are defined by the intended usage. However, most databanks will cover the domains of positional data (cultural, navigation, permit boundaries, etc.), geophysical data (seismic, gravity, and magnetics) , well data (drilling and completion operations, logging, casing, cores, mud, etc.), production data, transport facilities (pipelines and pumping stations), physical equipment, data carriers (tapes, documents) and facilities inventory data, interpreted data, and organisational data (access authorisation, authentication, organisational structure, partnerships, contracts, etc.). Figure 3 shows the use of a databank in the data life cycle of the asset exploitation process.

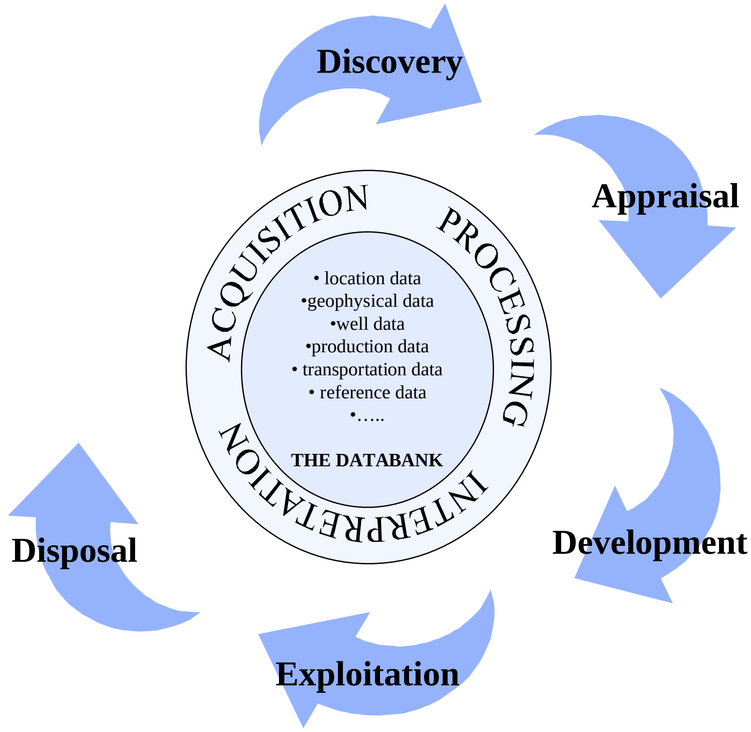

FIGURE 3: **THE DATABANK IN THE LIFECYCLE OF THE ASSET EXPLOITATION PROCESS**

The remainder of this paper describes the above steps in more detail.

Step 1 - Data Audit

The objective of the data audit step is to generate a reliable catalogue of all available E&P data. This is essentially the assessment of the "AS IS" state with a view to understanding the gap between this state and the "TO BE" state. This exercise is aimed at finding reasonably accurate answers to questions such as the location, state, quality, quantity, usability and importance of various data items. The data audit relies on techniques such as reports from existing databases (if any), user interviews, archive searches, and examination of statistically representative samples of data items (generally speaking, a complete audit of a physical data store would be prohibitive in cost).

The data audit generates information on the following :

- the data model structures of the existing systems. These could be simple indexing systems, or very sophisticated data models;
- semantics of particular data attributes and their possible time related evolution;
- physical condition of data;
- integrity of reference information (unique well names, seismic line names, etc.);
- relationships that enable one to associate data physically residing in different places;

- relative importance of data, and prioritisation of the subsequent capture exercise;
- various metrics such as rate of increase, frequency of modification, frequency of access for various data items.

By the end of the data audit step, one would generate sufficient knowledge to be able to establish the scope, timing, speed and estimated cost of the subsequent steps. A good data audit is an essential first step to improved data management.

Step 2 - Physical Upgrade

As stated earlier, many databank projects start out as physical data upgrade projects whose focus is to arrest physical deterioration of data, and improve access to data. The term physical upgrade in this context covers the areas of tape library transcription, document scanning, vectorising/digitising of maps, well logs and seismic sections when necessary, and is a first step in integrating these data items through better indexing.

Physical deterioration of data occurs due to a number of environmental, manufacturing, and general age related issues. For example, all magnetic media deteriorate with age. This process can be delayed by proper environmental control (such as temperature and humidity control), but it cannot be stopped. Paper and film also deteriorate with age. Additionally, some data types (such as Analogue, 21 track, etc.) become obsolete and therefore data needs to be moved to more modern media.

The physical upgrade programmes seek to move these items to new media through large scale transcription and scanning operations. These projects tend to be focused on optimised workflow since the number of items involved in any given project can run into millions of documents and tapes. An indexing operation is usually incorporated into these projects to ensure that the processed items are properly referenced and interconnected.

A desirable outcome of these projects is improved access achieved through compaction of the data library, and use of high speed storage devices that are available on the network. The results from these projects can also be moved to integrated databanks provided that the indexing information is compatible with the loading requirements of the target data model.

Step 3 - Logical Data Upgrade

The existing attribute data of an oil company typically resides in a number of places such as flat files, paper index systems, spreadsheets, operational databases, and application databases. The idea is to move all these data into an integrated data model environment based on POSC's Epicentre data model. The moving of data can be either on a "one off " basis (which means that the data source in question is now obsolete and will not be used for further data capture), or ongoing basis (which means that the database in question is an operational one and will continue to act as an ongoing data source).

The objectives of the logical data upgrade step are defined as follows:

- formally describe the source data stores in terms of contents and structure;
- identify the correspondence (mapping) of the source items to target entities in Epicentre;
- resolve issues such as data redundancy, multiple definition, missing relationships.

This step results in a formal definition of the processes to be utilised in moving data from source to target either on a one-off or ongoing basis. For ongoing exchanges, one needs to ensure that the scope of exchange is limited to only those items that need to be updated between two exchange sets.

Step 4 - Cleanup of Data and Transfer to Target

Even though step 3 has theoretically defined the exchange set and the source to target mappings, a considerable amount of work still needs to be carried out in cleanup and preparation of data for transfer to target. This need arises because of the lack of rigid control and appropriate use of reference values and existence rules in most legacy systems. Unfortunately, considerable time and money needs to be spent on correcting the data values to an acceptable level of consistency. This step would not catch all discipline related errors in the data, rather, it should be viewed as a minimal cleanup of data to ensure correct loading to the target. The actual attribute values may still require correction after the users are involved (step 5).

Step 5 - Connection to Applications

All the steps previously described offer considerable value for those involved in E&P data management. However, the impact for the end user is limited to better access to data of known quality. Although this is quite an impressive achievement, it falls short of the true end user requirement of accessing data directly from any application suite of choice. The prerequisite to true integration is the use of the Epicentre data model by all applications of interest. However, in the interim, the following approaches offer different levels of support for exchanging data with applications:

- provision of data in industry standard formats such as RODE, SEGY, RP66, LIS, UKOAA, etc. This approach is perhaps the simplest to implement but also limits the type and mode of exchange;
- provision of data to an established half link technology such as GEOSHARE. This approach enables one to build more intelligence into the exchange set. The disadvantage is that it is often quite difficult to keep encoders and decoders from different vendors synchronised. Therefore, it is quite typical for a half link to break following the release of a new version of software;
- use of proprietary “data publication layers”. Some companies have opted for developing “point to point” connections for moving data between their repository and

application environments. These connections stand a better chance of surviving software version changes, however, they are costly to build and maintain;

When POSC standards enjoy pervasive use sometime in the future, these interoperability issues will become less important. For the time being, we still have to rely on stopgap solutions.

Step 6 - Ongoing Data Acceptance Policies

When the legacy issues have been resolved, and the databank is ready for operation, one must ensure that ongoing data from partners, contractors, subsidiaries, and regional operating units are provided in well defined formats on acceptable media, and preferably in a manner which makes ongoing loading operations simpler than the legacy loading operations. To this end, policy documents have to be created to specify the footprint, format, extent, and estimated quality of each exchange set. It is envisaged that these policies would need periodical update to deal with new formats, data types, media, etc. Nevertheless, good data acceptance policies will go a long way towards ensuring that the databank contains reliable information.

Step 7 - Managing the Environment

The target hardware, communications and system management environment for an effective databank is largely dependent on the existing infrastructure, expected usage patterns and available budget.

It is clear that one can deploy a small databank with the stated aim of addressing the requirements of one asset team in a singular office. However, a more typical usage pattern is that of a multi-site, multi-continent corporate data store with LAN, Intranet and Internet access. Typical databanks of the very near future will contain hundreds of terabytes in robotically accessible nearline systems, with the ability to routinely deliver data anywhere in the world. This sort of power and flexibility brings up many issues related to infrastructure management, access authorisation, and authentication, data security, disaster recovery and general user support. Although out of the scope of this paper, these issues obviously have a large impact on the eventual success of a databank project.

Summary

This paper is an attempt to broadly describe the components of an integrated E&P databank. Careful planning of all the steps described above is a necessary, but not sufficient, condition of the construction of a successful databank. A major additional requirement is to promote the merits of good data management practices throughout the organisation, and encourage all stakeholders to spend the additional time necessary to ensure data integrity and quality, because ultimately, a databank is only as good as the data it holds.