AAPG International Conference & Exhibition '94
August 21-24, 1994, Kuala Lumpur, Malaysia

The computer revolution — implications for data management

MICHAEL J. WILTSHIRE

Wiltshire Geological Services
13 St. Andrews Avenue, Mt. Osmond SA 5064
Adelaide, Australia

Abstract: Trends towards reduced microchip size and power requirements coupled with increased processing capacity have profound implications for the future of the geosciences. Data acquisition, interpretation and storage capabilities are rapidly evolving, and data volumes now threaten to overwhelm potential users lacking appropriate data management tools.

Large volumes of digital data are now being generated. In the near future the management of information in our industry will become dominantly digital. Companies and individuals which master the new technology will thrive.

Much of the existing information base is analog, and substantial investment is required to convert the data to a concise, high quality digital form. The very detailed digital data sets now being acquired require quality control scrutiny and often preprocessing before being confidently and efficiently used by geoscientists. Both these tasks require skilled professional involvement. Careful management of both the existing knowledge base and of the new data now being accumulated is required.

No absolutely secure long term storage mechanism exists. Preservation of the oncoming avalanche of basic data cannot be assured. To successfully meet the challenge of information preservation, our industry must concentrate on efficiency and quality in our data preparation, storage and use. Mechanisms for funding the data conversion task, protecting the investment in data and in facilitating its routine use should become a priority. Attitudinal and organizational changes on the part of both corporations and Governments are needed, or the value of data will be unrealized and the data may eventually be lost.

INTRODUCTION

In the last 15 years we have seen continued development of computing power and parallel developments in miniaturization of componentry and computing systems. These developments have been accompanied by persistent and aggressive reductions in price of both components and systems. These trends show no signs of abating.

In many industries, we are now able to routinely monitor very large and diverse data sets, and to use these data as tools for interactive management and complex decision making. Good examples of the economic benefits are provided in both the drilling and geophysical exploration industries.

Historically, we have developed our technical expertise by observation and analysis of performance, and new technologies have consistently emerged from the innovative analysis of observational data. The development of the silicon chip, and the ongoing derivative development of computing power is an example.

We can expect the vastly improved ability to observe and analyse our environment to have a profound and unpredictable impact on the immediate technical evolution of mankind. Examples are the complete mapping of the human genetic code, with consequences (potentially both "good" and "bad" by our present moral codes) which cannot be foreseen or controlled.

The future development of computing power (both hardware and software) will be influenced if not controlled by positive feedback involving the availability of the existing level of computing technology and the wide availability of high quality observational and interpreted data in machine-readable form. With time, analog data will become increasingly irrelevant (because of access delays and costs, and lack of uniformity of display scale and style); we can expect that within a relatively short time leading edge explorationists will almost exclusively be using digital data.

Because of instability of the storage media, conservation of stored data requires regular monitoring and managerial interaction. To date, no secure and efficient long term storage mechanism

***Footnote** — This unconventional paper is included to stimulate consideration and hopefully debate of a part of our science which is becoming of crucial importance.*

for digital data has emerged.

Huge volumes of data are now routinely being generated. The long-term preservation and future use of these data, when considered with the large volume of historic analog data yet to be digitally converted, requires a significant and ongoing investment. Without such investment, data will eventually be lost.

COMPUTER HARDWARE DEVELOPMENT

Digital data manipulation and management first became generally affordable in the mid 1970s, due mainly to the development of the Digital PDP11 minicomputer. With the release of the DEC VAX in 1977, digital acquisition and processing of scientific data quickly became widely available. Subsequently demand for access to the power of such processors grew at an unexpected rate. It rapidly emerged that shared CPUs could not meet the ever-rising demand for desktop processing power.

Shared access to VAX-type processors remained relatively expensive, and data preparation and management costs were generally a minor project cost. Corporate information services departments administered hardware, software and data management, with technical users being isolated from such issues. The growing demand by technical users for access to more powerful processors, and more individual control of their computing environments, soon led to the next major development in technical computing.

During the mid 1980s, the dominant position of the minicomputer was attacked and quickly overwhelmed by development of workstations and desktop microcomputers. The underlying reasons were both technical (demand for more predictable and affordable access to CPU time) and social (reluctance to allow MIS managers to continue to control access to and use of information and computing resources).

By the late 1980s, desktop microcomputers having quite powerful CPUs were widespread in corporate environments, but little thought appears to have been given to the requirements for linking these machines. Individual manufacturers used differing CPUs, internal data formats, and hardware conventions, with no apparent concern for the needs of the user to share information. This led to the next major development theme, the adoption of standards and mechanisms for data sharing between machines — the networked computing environment.

Paralleling the need for improved ability to share data between machines (driven by the demands of the distributed desktop environment), was a need to more readily (i.e. naturally) manipulate data at the local level. Driven by rapidly evolving end-user expertise, this has led to the current developments in graphic user interfaces and visualization software.

The requirement for robust networks with high bandwidth and high quality screen display of detailed graphics has led to the development of powerful graphics oriented computers; these are now the technical workstations of choice in the exploration industries.

SOFTWARE DEVELOPMENT

Over the last 15 years at least, system and applications software development has lagged far behind the pace of hardware development. The reasons are complex, but include cultural and technical components.

The leading-edge life of a CPU platform is about 18 months, and the current generation of CPU technology is used to design the next. This ensures that small, independent and ambitious companies will constantly challenge the established hardware companies.

Chip manufacturers and system software developers have, to date, generally addressed fundamental hardware and software deficiencies and inefficiencies by use of faster CPU throughput and liberal applications of snake oil. Corporate fortunes rise rapidly and fall spectacularly, because the market is becoming increasingly expert at identifying companies which respect and value their customers.

As a consequence, the CPU manufacturers are preoccupied with hardware development and promotion, and software developers do not have (and have not for some years had) a stable and predictable base on which to develop efficient system and applications software.

Until system software provides users with elegant, efficient and robust tools for exploitation of the power inherent in the current generation of CISC and RISC processors (whatever they might be), applications software development will continue to lag far behind chip hardware development.

A more subtle influence on applications software development is the lack of large, verified, technical data sets in a widely available and cheap form. Without access to the large technical information base that exists in almost all industries, and without access to efficient and appropriate software tools, experts within each industry are severely handicapped in their efforts to develop new and better software.

Instead, commercial software tends to be developed by professional programmers, with little knowledge of or regard for what the end user actually wants. The wide availability of cheap RAM and disk storage, and the ever-increasing CPU throughput has resulted in a generation of large, inefficient and often unusable applications software. The current fashion for regular software "upgrades", sold to the existing client base, is evidence of the problem.

The Unix community is split between user/ developers of often powerful and appropriate software which is distributed at very low prices on the Internet, and commercial software (often migrated from the DOS environment) sold at inflated prices. Unfortunately, the user/developer software is limited by the Unix academic culture, which consistently discourages the new user by an arcane and utterly unnecessarily complex user interface.

AN AVALANCHE OF DATA

The wide application, increasing power and decreasing physical size of monitoring devices incorporating embedded CPUs has resulted in an enormous increase in our ability to observe and record our environment. Such methods have led to an avalanche of detailed technical data now being recorded. Development of the ability to systematically process very large and complex data sets then results in imaginative applications of the data.

Visualization software has become essential in many technical fields. It is simply not possible to use much of these data in the traditional manner — because of both the volume and immensely detailed nature of the data. Visualization software is the only way to handle such data in an effective manner.

The quality and timeliness of technical interpretation then develops rapidly. Examples are the recent developments in high resolution aeromagnetic data acquisition and interpretation, and in 3-D seismic.

These techniques rely on acquisition of new data. There is generally at best poor integration of the pre-existing data into the new data set and interpretation.

The primary reason is that the older acquisition technology constrains both the form and quality of the data, which are considered inferior to the newly acquired data. There is also a need to ensure the older data is consistent in format with the new data; this often involves digital conversion and other preprocessing. Such preprocessing is expensive, and is often considered to compromise the data.

The end result is that large and potentially valuable historic data sets are often ignored in the cause of cost-effectiveness and objectivity. The quality of regional interpretations is often compromised by such factors; it is too expensive to acquire the regional coverage needed for basinwide appraisal using the new technology, and too difficult to incorporate data products of the older technologies.

In mature exploration areas, the argument is often made that the regional picture is by now well known, and the costs of acquisition and processing of detailed regional data sets cannot be justified. Yet in many areas, superior interpretations arising from high grade data sets can lead to significant new plays.

In the case of the detailed subsurface data set presented by well logs, these data cannot be duplicated by modern acquisition, and the historic data set far outweighs the modern data set in terms of volume and technical data content. Unless the analog data are converted to digital, and integrated (in terms of data format and integrity) with the modern digitally acquired data, within a short time the analog material will not be effectively used in the digital work environment.

If digitally acquired data are not appropriately archived and subsequently correctly managed, the digital data will become corrupt and ultimately will not be usable without considerable restoration effort.

The rapid recent growth in volumes of data being acquired carries a deeper philosophic problem. In summary, even if we were technically able to store and recover the subject data at low or no cost, the volume of basic data is such that most of it will never be subsequently used. If this is the case, then efforts to store and ensure easy access to the data are pointless, and will eventually be abandoned.

Intelligent digital conversion involves far more than scanning of text and graphics, and subsequent digital storage of the images. Compression of images, and conversion of text to ASCII files, is the minimum required. Vectorization and editing of images, and skilled condensation of text is the preferred action.

In the process of conversion to digital form, it is possible to greatly condense much data without significant loss of the meaning of the data. The output data are more accessible and systematically searchable, and far more readily stored than the original material.

It is probable that the most effective data management will come from skilled data consolidation before storage, with systematic abandonment of the basic data after a suitable period, perhaps geared around the frequency of access to such data.

In the process of skilled data consolidation, the basic data are verified, and problems of data quality can be corrected, or at least identified. If such work is done systematically, soon after initial data acquisition, we have a better chance of full recovery from any problems identified with the basic data.

MANAGEMENT OF DATA

For effective and ongoing use of older data (whether analog or digital at first acquisition), digital conversion and subsequent management of the digital archive are required. This is a requirement in many technical industries; there is nothing particularly special about the petroleum exploration industry

The costs of such work are considerable. From the viewpoint of a company wishing to use the data, the costs of conversion and management of a large archive are significant but definable. In the absence of a pool of publicly available digital data, the opportunity costs of delays in acquisition of the data in digital form are potentially greater than the costs of conversion, but are intangible.

This is clearly a situation where a pooled industry effort and sharing of costs is desirable, and has great potential benefits to all. The benefits to governments are even greater.

To establish an effective, widely available and affordable digital archival service requires, however, an acceptance that the value inherent in the information will only be realized by its application to the technical problem we are addressing, and that the best quality interpretations will only arise by the wide dissemination and use of the data.

The economic benefits of more efficient exploration and production will mean greater potential profits to both companies and governments. This scenario will only be achieved in an environment where information exchange (digital or not) is encouraged, and where economic benefits to individuals or governments arise from a more profitable and successful industry.

In short, digital data management and its application to the exploration task requires a cultural acceptance of an open society in which such data are freely available, or at least tradable without undue bureaucratic interference (financial or technical).

If the exchange of data is restricted by (for example) governmental licence requirements, an environment is created in which companies may elect to bypass officially sanctioned channels to data, either because of doubts as to the integrity of the data, or their ability to operate subsequently on their interpretations of the data.

In such a situation, unsanctioned trade of data between companies will eventually bypass the governmental restrictions, the official data archive will be inadequately funded and will eventually fall into disrepair. The end result will be that the government's ability to monitor and manage the industry will be limited because the companies' data resources will be more detailed, better maintained and more regional than that of the governments.

Geology knows no political boundaries, and international companies commonly operate on both sides of such boundaries. Where governmental control of a single geologic province is divided between two or more states or nations, unless the states are prepared to exchange the basic data, each state will soon find that their understanding of the basic geology and exploration consequences will not compare with that of the international companies.

When companies enter into negotiations with governments regarding exploration or exploitation conditions, companies are obliged to seek conditions which will maximize shareholder benefits (or potential benefits). In negotiations between government and companies, companies with superior information will be able to negotiate better terms than those lacking the information.

Societies which restrict the flow of information in the exploration industry, for whatever reasons, ultimately place themselves at a competitive disadvantage relative to the international companies and the more open, information based societies. International exploration companies concentrate where the conditions for cost-effective exploration are best. One element in the equation is the availability and quality of the existing data resource.

For reasons such as these, effective data management is an integral part of an efficient exploration industry. As the industry converts to digital data management, governments worldwide must address the question of the conversion and ongoing management of their data resource.

A number of societies currently seeking to introduce western methods and management to their exploration industries are attempting to reconcile past attitudes to the free flow of information in their societies with the needs of the future. The changes are not coming easily.

AN HISTORIC ANALOGUE

The advent of powerful technology inevitably leads to significant social change. It is instructive to compare the potential impact of the widespread penetration of our societies by personal computers with a similar event in human history.

Consider the invention of the printing press, and the impact that invention has since had on mankind. There are many parallels with the computer "revolution" we are now undergoing.

Prior to the invention of the press, power in society was primarily held by those controlling access to knowledge, and since dissemination of that knowledge on a large scale was not possible, power remained concentrated in relatively few hands. (Modern examples of immense social change resulting from dissemination of information in society are the changes recently seen in the Soviet Union, and in South Africa. Many states currently seek to control the flow information into their societies.)

With the invention of the press, a mechanism was presented for cheap distribution of information. The process involved the conversion of the previous knowledge base to the new form, the development of an education system, and the eventual distribution of knowledge to a much larger sector of the population than had been previously possible.

The process took some time to gain momentum, but, once started, the consequences were both unpredictable and unstoppable.

The key element producing social change through the press was the information conveyed by the medium. Assembly of data and its presentation in the new form was a fundamental prerequisite for the changes that followed.

Viewed in this context, it can be seen that the task of data conversion and preparation for the digital future has far greater consequences than the short term exploration applications of the data. The presence of a large knowledge base in machine-readable form is a necessary condition for the evolution of computer technology, and consequently of society.

Those of us involved in the information industry thus have a responsibility which extends beyond the immediate applications of the data. We must endeavour to present accurate, cost-effective data, and ensure that it is widely distributed and used. The future lies with those societies that recognize and embrace the opportunity presented.

CONCLUSIONS

The tide of technologic change now in train, arising from the development of the microchip, will progressively overwhelm all aspects of our industry. Digital data conversion and requirements for subsequent data management are inevitable consequences of the microcomputer revolution.

Because of costs and management requirements, the effort is best conducted by a single group acting in collaboration with the host government. In a buoyant exploration industrial climate, self-funding of the effort is possible by use of modest private seed capital, and a high rate of reinvestment from cash flow.

To date, at least five countries have addressed the question of digital conversion and subsequent management of their well log data. These are Britain, Norway, Canada, Australia and New Zealand. One has achieved this through a governmental agency, one by industry in collaboration with government, and three by private efforts.

All seem to work effectively to the satisfaction of their domestic industries; various private efforts are now underway to transfer the technical skills and benefits to other nations.

Such efforts must be commercially viable, but must also deliver the data and management services on terms acceptable to the host nations. Because of the great cultural, economic and prospectivity differences between various nations, negotiating mutually acceptable terms for such efforts is not a simple matter.

The technologic revolution we are involved in will inevitably lead to great changes, both in our industry and in society. The question we must all address is not whether these changes are avoidable, but how best we can all adapt to and benefit from the opportunities presented.

ACKNOWLEDGEMENTS

I wish to thank Dinah Pantic and Andrew Deer for constructive comments on the text.

SUGGESTED REFERENCES

General references:

BYTE — Monthly computing newsmagazine — McGraw Hill Publishing, Peterborough, NH. All issues since 1980 contain relevant material.

VARIOUS NEWS MEDIA REPORTS (print and television) over the past five years, particularly in *The Australian, New Scientist* and *AAPG Explorer.*

GENERAL SOCIAL REFERENCES include writings of Bertrand Russel, Marshall McLuhan, Gore Vidal, Alvin Toffler and John Naisbit.

More detailed reference to the impact of computing technology on the earth sciences:

BROWN, ALISTAIR D., 1991. Interpretation of 3-Dimensional Seismic Data — *AAPG Memoir 42,* AAPG, Tulsa, Oklahoma

THE LOG ANALYST — SPWLA, Houston, Texas. Bi-monthly *Journal of the Society of Professional Well Log Analysts.* Many articles, particularly over the period 1985 to present.

Manuscript received 24 August, 1994