Explor97 Master Page ◀ GOTO ▶ Author Index
Explor97 Contents ◀ ▶ Section Contents
Previous Paper ◀

An Earth Science Information / Process (ESIP) Model and the Data Processing and Analysis (DPA) Class of Geoscience Software Systems

Hollyer, G.M.[1], Dobush, T.M.[1], and MacLeod, I.N.[1]

1. Geosoft Inc., Toronto, Ontario, Canada

ABSTRACT

Decades ago, as the world was preparing the groundwork for the computer revolution, the term "Information Age" was introduced. Information Age "thinking" brought entirely new ways of looking at the human problem-solving process—one of which was a new definition of the terms "data", "information" and "knowledge" to reflect the human and computer-based processes applied to the basic building blocks of "data" or "signals".

Although these definitions have since become an accepted part of our scientific and societal lexicons, their significance as revealed years ago has not diminished. These terms still provide a means of conceiving and organizing the basic activities in which we, for example as geoscientists, are engaged. In this paper, we introduce a model referred to as the Earth Science Information/Process (ESIP) model. This model is intended to provide a general framework for viewing the basic geoscience problem-solving process and the relationship of data, information and knowledge throughout this process.

This paper initially defines the ESIP model in terms of its basic process and information components. It relates these components according to their implications in terms of the geoscientist's problem-solving process, tasks and tools. We then apply the model to examine a specific class of software-based problem-solving tools, called Data Processing and Analysis (DPA) systems—systems which provide the capabilities required to manipulate or develop specific components of the ESIP. We also provide both ESIP and functional descriptions of the DPA software class.

In summarizing our results, we are proposing the ESIP model as a conceptual organizing structure for:

- *Understanding the geoscientist's problem-solving process and the milestones (components) along the process pathway (creation of information, development of knowledge and decision-making).*
- *Understanding the role and implementation of supporting PC-based software systems through which the geoscientist manipulates data and information to develop specific knowledge and make key decisions.*
- *Understanding the implications of different software tools for performing specific roles. Although not detailed here, the ESIP model provides a means of examining different types of software systems (such as Geographic Information Systems (GIS)) and determining the optimal combination of Geoscience "Authoring" (DPA), "Library-Retrieval" (GIS) and other role-fulfilling software.*
- *Understanding how geoscientists function. A key result may be in helping to implement organizational structures and software tools that are truly matched to the specific tasks geoscientists perform and to their fundamental needs as earth science professionals.*

In "Proceedings of Exploration 97: Fourth Decennial International Conference on Mineral Exploration" *edited by* A.G. Gubins, 1997, p. 185–188

INTRODUCTION

The ESIP model is proposed as a starting point for understanding the general earth science problem-solving process and the DPA class of software tools on which this paper is based. Before proceeding, we pause to define the ESIP model and illustrate the model's role in relating data, information and knowledge to the basic activities in the geoscientist's decision-making process.

As shown in Figure 1, the ESIP model uses a pyramid to express the relationship between data, information and knowledge (defined as "primary components"), and the acquisition and decision components.

At the base of the pyramid is a measurement or sample. The acquisition component considers all relevant constituents related to data gathering, such as survey design, quality control, and physical or electronic collection of measurements or samples.

Data comprise the next level in the hierarchy. Stated most simply, data represent unprocessed observations. For instance, a data item could refer to a specific geologic observation, a geophysical measurement made in an airborne acquisition system, a geochemical assay value, a sample location or an observation received at a satellite sensor in orbit around the earth.

The next level in the hierarchy is information. Information contrasts with data in that it represents an order or pattern recognized in the data. Essentially, raw data has been transformed through some process (human or computer-based). Examples in the earth sciences are the following:

- geologic lithology map constructed from related geologic observations,
- filtered, leveled or gridded geophysical data,
- geochemical observations that have been manipulated statistically,
- gridded topographic data, and
- filtered, leveled or gridded satellite images.

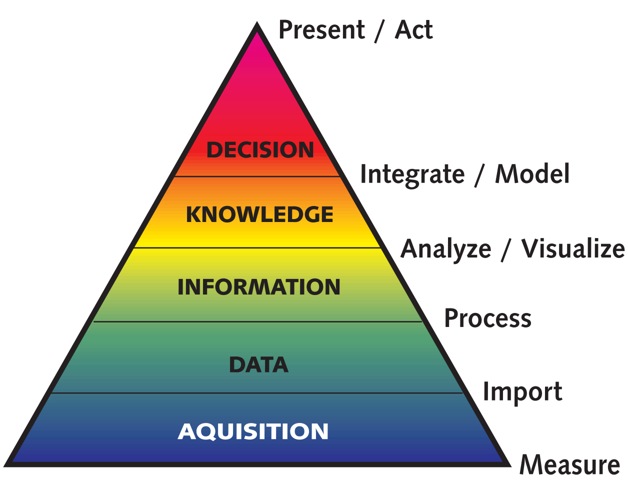

Figure 1: *The Earth Science Information/Process Model (ESIP) is a representation that describes the life cycle of data as raw data is transformed into information and knowledge. When knowledge is realized, data completes its primary life cycle of usefulness—and a decision is made. Archived data may be retrieved at a later date for further evaluation (secondary life-cycle).*

The next level is knowledge. Knowledge represents organized information. Knowledge is the sum total of all experiences, which combines the geoscientist's practical experiences and formal educational training, and the analysis and visualization of data or information. One recent example in engineering literature (Ferguson, 1992) distinguishes this type of experiential, combined knowledge as the deeper workings of the "mind's eye." This view contrasts with the task of "visualization," which can be thought of as simply one of many windows to the "mind's eye".

At the top of the hierarchy lie decisions. The final outcome in any earth science investigation is to make a conclusion and act on this conclusion (or recommend an action). In mineral or petroleum exploration, the most common decision is whether to drill at a certain location. In environmental engineering, the decision may be to recommend a more detailed survey prior to remediation on a site where a specific contamination problem is recognized. In Unexploded Ordnance (UXO) detection, the decision may be to remediate a subset of the detected targets to specific depth on a certain site.

Implications of the ESIP model

As embodied in the ESIP model, we can further recognize certain characteristics that reflect both the pyramidal representation and the levels of organization within.

- The data level is the most abundant component in the system followed by information and knowledge.
- Data must progress through one or more transformations to become information.
- The data transformation is a focusing process—in which scientific and economic value is increased through the manipulation of large volumes of data into information, knowledge and finally, decisions.
- Decisions are ultimately based on data—when we break, forget or consciously choose to ignore this connection, we may be subject to a loss of "data context" and/or "data intimacy".

Relating the ESIP model to the geoscientist's specific tasks and tools

The ESIP model has a variety of implications. Firstly, the model provides a way of understanding the geoscientist's "point-of-entry". Each individual relies on data, information and knowledge in varying degrees and the model can help determine dominant components.

For example, as geoscientists navigate the hierarchy of their own organizations, they tend to become more isolated from data and rely more on information, knowledge and decision-making skills. On the other hand, a less senior geoscientist may rely more on intimate knowledge of data, collecting and organizing data to a point at which resulting information may be passed to a senior geoscientist. Alternately, a geoscientist involved in an integrated organization may be required to navigate all levels of the ESIP hierarchy and work with all primary components to reach a final decision.

The identification of ESIP involvement, in turn, has significant outcomes in terms of evaluating and using specific classes of software tools. This exercise is non-trivial since many geoscientists have come to rely on software as a key resource. If geoscientists primarily work with data and information, they require different tools or different means of interacting with tools than geoscientists who function at the decision-making level.

A second implication of the ESIP model relates to the capabilities of today's generation of software tools. Today's emerging generation of PC-based DPA systems are increasingly using interactive, visual implementations to add informational value and to help build knowledge. This approach mirrors the "mind's eye" process described earlier in that these new implementations enable the user to interactively combine experience with "point-and-shoot" processing, analysis and visualization functions.

In the remainder of this paper, we look at the class of software systems referred to as Data Processing and Analysis (DPA) and see how these systems apply the ESIP model to assist geoscientists in working with data, information and knowledge. In addition, we look at the way in which DPA systems help geoscientists augment existing knowledge with entirely new knowledge generated from data and information.

OVERVIEW OF THE DATA PROCESSING AND ANALYSIS SYSTEM (DPA)

The Data Processing and Analysis (DPA) system is a class of software that focuses explicitly on the data-information-knowledge spectrum of the ESIP model. DPA systems trace their evolution to exponentially increasing data and information volumes and to the development of PC operating systems and hardware capable of handling high-volume data.

These types of system are routinely used to assist geoscientists in addressing problems related to data and information—such as removing cultural and geologic noise that obscures true geologic "signal" and manipulating high-volume data sets. Common DPA applications include processing, handling and management of data and information from mineral and petroleum exploration, environmental engineering, unexploded ordnance detection and a variety of other applications.

ESIP description of DPA systems

As shown in Figure 2, DPA systems are designed specifically for the geoscientist who works with data, information and knowledge. The user can enter the system via any of these primary components.

The orientation of DPA is process-based—an approach that is defined in terms of the Data Stream concept. The Data Stream comprises five major components:

- *Import.* Facilitates access to data and information.
- *Process.* Provides capabilities for converting raw data to information.
- *Analyze/Model.* Provides capabilities for converting processed data into geologically meaningful quantities and models.
- *Visualize/Integrate.* Provides capabilities for applying knowledge through the geoscientist's experience in extracting information from visual representations of data and information.
- *Present/Act.* Provides presentation facilities for creating maps and tables on which decisions are reached and actions recommended.

In essence, DPA systems can be thought of as earth science "Authoring" systems—systems in which the geoscientist has the ability to start with raw data, process it to eliminate problems evident in the data and

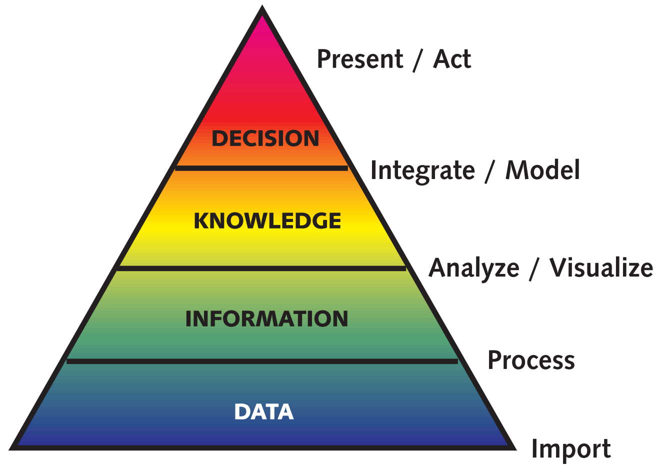

Figure 2: *The DPA representation incorporates both the ESIP model and the subset of activities performed within a DPA software environment. The five activities to the right of the figure constitute the DPA Data Stream in which raw data is transformed to information, knowledge and decisions through specific computer-assisted tasks.*

ultimately transform the data into information and knowledge. The geoscientist functions as the author, writing the "book" from start to finish and resolving one or more problems during the process.

From a problem-solving viewpoint, systems that allow experts to become "authors", such as DPA, have a number of practical implications.

Firstly, the original and processed data remain in the system so that the user always retains "data context" (i.e., contact with the data, resulting information and the processes that have been applied to the data). Loss of data context is increasingly problematic in a complex working environment in which volumes of data and information are rapidly increasing. Moreover, increasing time pressures have increased reliance on third-party data—a trend that has the potential to isolate us from data and the processes applied to the data.

Secondly, DPA systems provide unique means of connecting data and information to the geoscientist's "mind's eye". For instance, real-time dynamic linking mechanisms now enable the user to apply new visual "querying" methodologies and immediately relate numerical (0-D) representations of raw and processed data to related information, such as 1-D data representations (profiles, histograms, scattergrams) and 2-D data representations (maps).

Thirdly, these systems function as "experiential" systems. Because users can process the data from start to finish, they have maximum exposure to data. This provides a wide window of opportunity in which to increase the understanding of the data and ultimately, to gain knowledge about the specific problem at hand.

Functional description of DPA systems

Functionally, DPA systems can be defined by specific capability and related position in the data stream:

- *Store located (i.e., georeferenced) earth science data (Import).* Data can be accessed in ASCII or binary form and stored internally in the system.

- *Store located earth science information (Import).* Forms include vector- and raster-based grids, images and maps.
- *Data manipulation platform (Process).* Platform functions as an integrated data processing "engine". A Graphical User Interface (GUI) and integrated database may provide a core environment for manipulating data and information.
- *"Vectorized" database architecture (Process).* Columnar architecture (as opposed to standard relational database architecture) and related data handling mechanisms enable efficient high-volume data processing.
- *Basic and advanced data manipulation (Process).* Basic routines may include noise reduction (filtering), math expression and other functions for enhancing the quality of raw data. Advanced routines may include frequency domain filtering, gridding and advanced grid manipulation.
- *Information processing (Analyze/Model)*. Routines help convert processed information to knowledge. Examples include kriging, and forward/inverse data modeling.
- *Process querying (Process and Analyze/Model).* As the users progress along the data stream, they can instantly view a summary of the processes applied to data and information.
- *Conventional querying (Analyze/Model).* When working with spreadsheet (0-D) data, the user can extract all values greater than a certain threshold within a column and place the resulting values in a new column. These values can then be processed further or placed on a map.
- *Data and information merging (Integrate/Visualize).* Graphical User Interfaces (GUI) enable geoscientists to simultaneously work with *original and processed data* in spreadsheet (0-D), profile (1-D) and map (2-D) formats.
- *Visual querying (Integrate/Visualize).* "Dynamic links" connect data in 0-D, 1-D and 2-D representations or views. When a location in a spreadsheet, profile or map is selected, cursors automatically locate the corresponding data in other views.
- *Imaging (Integrate/Visualize).* Users build composite 2-D maps by combining raster- and vector-based data and information. Maps can then be manipulated interactively to extract subtle features in data and grids (information).
- *Mapping (Present/Act).* Users can employ map editing and output tools to prepare final presentations for further analysis and decision-making.

From a functional perspective, the DPA system can be summarized as a software environment in which unprocessed data is imported into vectorized databases for access, processing and analysis, and in which processed data is stored in vector-/ raster-based maps. Maps are intelligent entities that remain linked to original data and that provide a powerful means of quickly analyzing, visualizing and presenting large volumes of data.

CONCLUSIONS

The Earth Science Information / Process (ESIP) model provides a general framework for understanding the geoscientist's problem-solving process and the milestones along the process pathway (creation of information, development of knowledge and decision-making).

In addition, the ESIP model provides a platform for understandi the role and implementation of supporting PC-based software systems through which the geoscientist manipulates data and information to develop specific knowledge and make key decisions. This process was examined here through discussion of a specific class of system called the Data Processing and Analysis (DPA) system. The ESIP model provides a starting point for conceptualizing some of the basic characteristics of DPA systems:

- Point of attack—DPA starts with data, and/or information, depending onthe geoscientist's needs and organizational role.
- Data-driven— "Vectorized" database structure is constructed to handle high-volume data as it is transformed from raw data to model, image or report.
- Process-driven—Data stream summarizes the activities that span the model. Data stream is not necessarily linear since the user can repeat tasks or reverse the order of tasks for a particular application
- Link data, information and the knowledge accessed via the "mind's eye" —Mechanisms such as process querying, conventional querying and visual querying enable instantaneous direct linking to the geoscientist's pool of combined experience and professional training.
- Value-based—Progressively adds value to original data by applying processing, analysis, integration, visualization and by reporting processing "events".

The ESIP model also provides a benchmark for understanding the implications of different software tools for geoscientists. Although not examined explicitly in this paper, the geoscientist can use this model to evaluate other software systems. An example is the Geographic Information System (GIS) which operates primarily at the information, knowledge and decision-making levels of the hierarchy. Common GIS functions are to provide access to information and develop knowledge through standard querying and mapping processes. These systems generally function more as "Library" systems than "Authoring" systems.

Finally, in an environment in which geoscientists are being asked to re-invent their roles to deal with increasing data, information and competitive pressures, the ESIP model may help us understand how geoscientists function. A key result may be the development of organizational structures and tools (such as software systems) that are truly matched to the specific tasks that geoscientists perform and to their fundamental needs as earth science professionals.

REFERENCES

Ferguson, E., 1992, Engineering and the Mind's Eye, Cambridge, MA: The MIT Press, 194 pp.