Technical Paper

Advances in Commercial Software for High-Volume Geophysical Data

Ian N. MacLeod and Greg M. Hollyer

Introduction

The instant that a geophysical measurement is made inside a modern geophysical sensor is converted into a digital value and stored inside a computer. It will live its entire useful life in digital form – occasionally (and ultimately) achieving its full potential when displayed as a visual image that communicates geologic significance to the eyes and thinking brain of a human interpreter.

The trail that that geophysical data value will follow is remarkable. In our attempt to improve its value and character, it will be attenuated, shifted, amplified, decimated and bred with other data to produce offspring. All of these improvements (and some injustices) are carried out by computer software. While this integral role of software seems obvious, it is important to remember that we are completely dependent on the capabilities of our software tools to deliver to us the information value of our data.

In this paper, we focus on the unique requirements of airborne data and airborne data processing, and how we have addressed these requirements in a new commercial software system called Oasis montaj. We will look at how the application and development of software has changed in the past 10 years, and how this new generation of software addresses problems that are specific to modern airborne data processing. Key issues are databases and data storage; how we deliver new functionality; and links – links between the underlying data and what we see and links with the processes that have created what we see.

A Decade of Change in Airborne Data Processing

We have seen a number of important trends in the development and application of airborne processing software over the past 10 years:

1. An order of magnitude increase in the volume of data collected for similar size projects.
2. Steadily increasing need to integrate different sources of digital data as part of the interpretation process, with the result that interpreters today are more "digitally connected" to their data.
3. The movement of applications from "head-office" workstations and mini-computers to field and exploration office PC based workstations.
4. The replacement of proprietary software with industry-standard commercial software.

Ten years ago, there were few commercial software tools designed to work with airborne data and the majority of software was developed in-house by those who handled, processed and applied the data.

This reliance on proprietary software led to a wide variety of ways to handle, process, visualise and interpret airborne data. Although there was much room for innovation, proprietary software presented a barrier to those who did not have access to those software tools, or lacked the computers, skills and time required to develop their own software. Interpreters generally worked with fixed map or image products delivered by data processors. They had little or no contact with the original data or the processes that created the maps.

In addition, proprietary software resulted in significant duplicated effort as different groups developed tools that implemented effectively the same functionality. Moreover, in-house resources were devoted to the support and maintenance of these systems rather than the use of the systems.

Commercial software for general geophysical data processing has evolved into the common industry standard applications that we see today. These developments have been due in large part to the availability and power of microcomputers, which have led to a sufficient number of potential users for commercial software companies to survive and even thrive. With the acceptance of Windows NT as a PC-class workstation platform, commercial developers have now turned their attention to airborne data processing.

The current availability of commercial PC software means that anyone with the resources to acquire a commercial software license can process and work with digital airborne data. It allows data processors and data users to spend more effort interpreting data and even develop new and innovative techniques that build on existing commercial systems.

New Solutions to Data Processing Problems

In the design of Oasis montaj, Geosoft has addressed the following three software technology challenges:

1. Data volume and the unique character of airborne data.
2. Functionality and the development of a model to allow the addition of new processing functions.
3. The creation of links between final maps and the original data, and the processes that have been applied to the data.

Airborne Data Sets are Large

A 10,000 line km airborne magnetic and radiometric survey can generate 270Mb of raw data – a size that can double or triple as the data is processed. Therefore, modern software systems must be able to comfortably handle data sets of this size and possibly much larger.

The issue of data volume is primarily related to the capabilities of the underlying operating system and hardware. Experience has shown that 16-bit PC operating systems (MS-DOS and Windows 3.1), are simply not up to the task of handling very large data sets. Memory addressing limitations in these systems requires considerable programming effort to overcome, and there has been a corresponding limitation in the capabilities of the underlying hardware (disk size, memory and performance).

A solution to these problems has emerged with the development of a widely accepted 32-bit operating system for the PC architecture - Windows NT. The power of Windows NT has led to the availability of new “workstation” class PCs, which are equipped with 5, 10 or even 20Gb disk drives and 64 to 128 Mb of RAM. In addition, CPU, graphics and mother board performance has improved significantly as hardware manufacturers strive to keep pace with the demands of new software designed to take advantage of this “workstation” PC environment.

Airborne Data is Different

Airborne data is quite different from many other types of geographical data. Airborne data is collected at a high sample rate along long lines, and many channels of information are often recorded at different sample intervals. For example, positioning data and radiometric data are normally sampled once per second, magnetic data is typically sampled ten times per second, and EM data may even be sampled 20 times per second. By extension, the data storage structure in a computer should be able to accommodate these differences.

Conventional Database Approach

At this point, it is worth looking at conventional relational database technology (SQL, X-Base, Access, etc.) to evaluate its effectiveness for storing and working with airborne data. It would seem to be a good idea to store geophysical data in relational databases because they already exist, they are mature in their stability and functionality, and they offer numerous tools for maintaining the data and data structure.

Relational databases store data in a structure based on tables, records and fields. Figure 1, shows how we might store simple airborne magnetic data in a conventional database table. A table contains many records, one for each measurement location, and each record contains many fields that contain the data values.

However, there are a number of problems with this data structure. Relational tables are most efficiently accessed by record. Therefore, if you want to access a single field in a record, you must read the entire record and extract the field of interest. Another problem arises when you want to change the record structure, for instance, by adding or removing a field. Although relational tables allow records to be efficiently added, inserted, changed or deleted, adding or removing a field is computationally intensive.

CONVENTIONAL DATABASE TABLE

Fields

Records

line	fid	x	y	mag	height
100	1	10517.1	8013.2	56600.4	120.0554
100	2	10517.2	7977.5	56598.7	120.8992
100	3	10516.8	7940.3	56595.1	121.7662
*	*	*	*	*	*
200	2961	9585.8	7998.6	56720.6	122.0075
200	2962	9585.6	7956.1	56711.7	122.8997

Figure 1: This figure illustrates the way that a simple magnetic survey data set is stored in a conventional database table. The table contains many "records", and each record contains a fixed set of data "fields".

To see how these limitations would effect airborne data processing, consider the example of applying a convolution filter to the "mag" field shown in Figure 1 and saving the result in a "filter" field. To perform these operations, the software must:

1. Read each record, extract the "mag" field, and store it in an array.
2. Apply the filter to create a new filtered array
3. Call a database function to re-structure the table to add a "filter" field. The database software will effectively create a new table with the desired structure; read each record of the old table and fill in and write the record into the new table, then delete the old table. This can be a slow process.
4. Write the filter result into the new field by reading each record, move the next filter array element data into the new "filter" field, and write the record.

This read-process-write sequence is repeated many tens or hundreds of times as data is reduced from raw field data to final form, and it must efficient – especially considering the volume of data involved. The fact that many processes also require lines to be processed separately further complicates the task.

Another problem arises working with data sampled at different sample intervals. The conventional table structure requires all data in a table to be sampled at the same interval, which means that the system either interpolates fields and stores redundant information or creates separate tables to store different information. The latter option adds to the complexity of maintaining relationships between the tables.

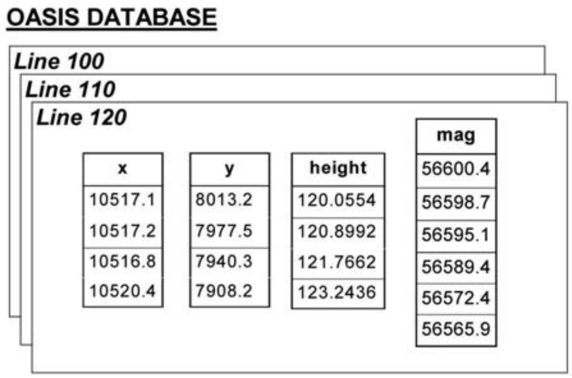

Figure 2: THE OASIS DATABASE STORES AIRBORNE DATA AS A COLLECTION OF "LINES". EACH LINE MAY CONTAIN ANY NUMBER OF "CHANNELS", AND EACH CHANNEL CONTAINS AN ARRAY OF DATA "ELEMENTS". CHANNEL ARRAYS MAY HAVE DIFFERENT SAMPLE INTERVALS.

The Oasis Database

Oasis databases organise data as lines, channels and elements as illustrated in Figure 2. The database may have any number of survey lines; each survey line includes any number of channels, and each channel stores an array of data elements of any type. Channels are stored separate from each other, and may have different start points and sample intervals. The Oasis database is designed to be able to quickly create and delete lines and channels, and all data in a channel can be treated as a unit[†].

In contrast to the conventional database approach, the application of a filter in Oasis is as follows:

1. Read the "mag" channel data from a line into an array. Because the data is stored as an array, this step simply moves the array data from disk into memory.
2. Apply the filter to create a new filtered array.
3. Create a new channel called "filter". This only involves creating a new databse channel symbol - no data is read or written.
4. Write the filter array to the new "filter" channel. In comparison with the relational database example, there is no wasted effort in this process and a maximum efficiency is achieved.

Working With Data

As related in the introduction to this paper, airborne data passes through many separate steps on its way to its final destination on a map or image. From a database technology viewpoint, airborne data is considered extremely "volatile". This means that it should be stored to allow changes to be made as efficiently as possible. The example scenario described previously illustrates the importance of an efficiently designed database.

It is also important to provide visual feedback to the data processor so that the results of processing steps can be seen and evaluated. In Oasis montaj, we take advantage of the Windows graphical user interface to provide an effective data processing environment.

As shown in figure 3, montaj presents a user with a project "workspace", within which all processing and visualization takes place. Within a workspace, a spreadsheet window is used to display the numerical contents of an airborne database. The spreadsheet can be thought of as a view of the data contained in a database. The user selects which channels are viewed and how to display those channels. Because different data channels may be sampled at different intervals, the system needs to do something in order to display channels together in the same spreadsheet view. In these situations, montaj resamples channels to the lowest common denominator and displays re-sampled channels with a red line under the channel name to indicate that this has happened. Re-sampling is only done for the purpose of displaying the data, so the original data in the database remains unchanged.

† The concept of storing data as separate channel arrays was developed by Scott Hogg and Ray Whitten, who used this idea in the development of the Aerodat database in the early 1980s.

From this interface, the user can directly modify any information in the database, either by running a process, or by changing the data directly in the spreadsheet view. As you would expect in a spreadsheet, you can replace the contents of a channel with a mathematical expression that uses any database channels as variables.

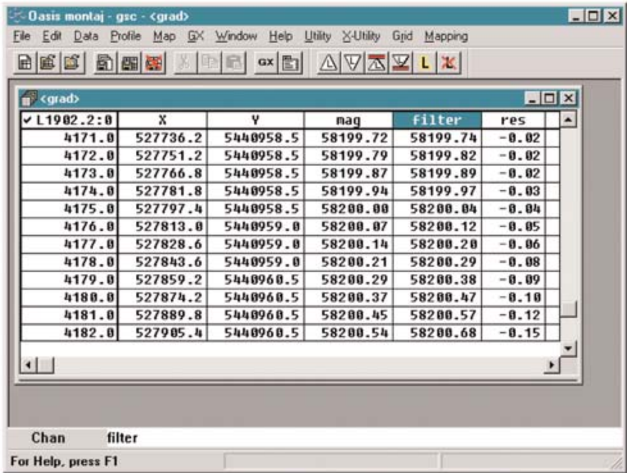

Figure 3: An Oasis montaj project workspace showing an open database displayed in a spreadsheet. The survey line number (type number.version:flight) is shown in the top left corner and the line fiducial is in column below the line number. Channels are automatically resampled in order to be displayed with other channels.

We like to think of the spreadsheet view of the data as the 0 dimensional view - this is the real data, point by point. With airborne data it is important to be able to view the data in profile form. Figure 4 shows the profile view, or 1-dimensional view of some channels in profile windows beneath the spreadsheet. The profile cursor and spreadsheet mark are linked so that the user can relate the profile back to the original data.

Finally, 2-dimensional views of the data are provided through map windows, which may contain any 2-D graphical information. Figure 5 shows two maps, one containing a total field image and a second showing a contour map. Maps also contain a crosshair cursor that maintains a link with the profile view and the spreadsheet view of the data. The processor is able to visualize and relate her data in 0 dimensions (spreadsheet), 1 dimension (profile) and two dimensions (maps). This gives a natural feedback for making decisions about processing parameters.

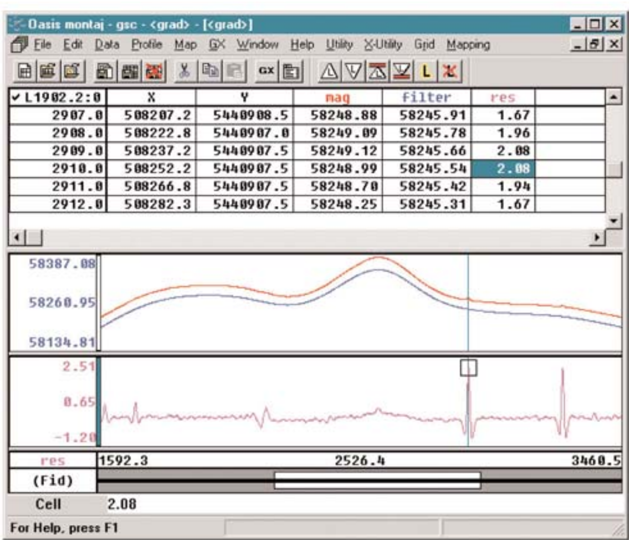

Figure 4: The spreadsheet represents the zerodimensional view of the data and the profiles provide a one-dimensional view.

While many processes require visual feedback, there are also many that are best run in "batch" without user interaction. To accommodate these needs, the system is able to execute processing scripts either from the Graphical User Interface (GUI), or from a separate script processing program.

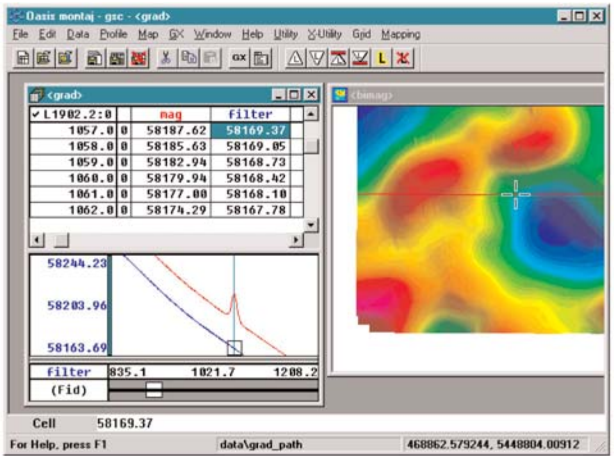

Figure 5: THE TWO-DIMENSIONAL VIEW IS PROVIDED THROUGH A MAP WINDOWS. THE CURSORS IN THE SPREADSHEET, PROFILE AND MAP WINDOWS ARE ALL TIED TO THE SAME LOCATION.

Adding Functionality

After data is imported from an acquisition system onto an Oasis database, the geophysicist applies numerous processes as part of the compilation process. For magnetic data, this includes systematic survey compensations, de-spiking, pre-filtering, base station corrections, leveling, IGRF removal and postfiltering. There are also many interpretive processes that may need to be applied to the final data, such as vertical continuations, derivative calculations, polereductions and special purpose filters.

One of the special challenges in designing Oasis was to find a way to easily add new functionality and provide users with the ability to add their own functions. To do this we developed the concept of Geosoft eXecutables, or GXs.

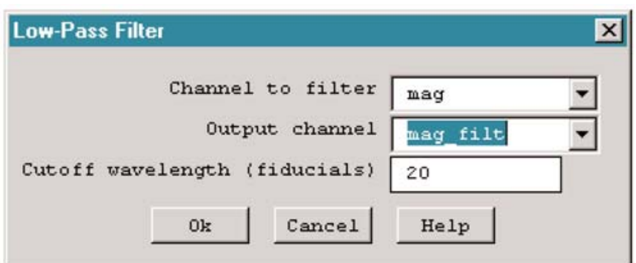

Figure 6: SAMPLE CONTROL DIALOGUE PRESENTED BY A LOW-PASS FILTER GX. ALL PROCESSING FUNCTIONALITY IN OASIS IS CARRIED OUT BY GXS, WHICH CAN BE DEVELOPED BY END-USERS.

As shown in Figure 6, a GX is a “program” that can be used to perform a specific task within the Oasis montaj environment. GXs are written using the GX Developer and the GX language, which is a simple C-like programming language that gives a GX application developer access to the capabilities in the base system (such as expressions, filters, FFTs, etc.). GX developers also have control over various graphical user interface (GUI) elements (menu bars, dialog boxes, etc.) so that new functions can be well integrated into the end-user environment.

All of the application tools delivered by Geosoft are provided through suites of GXs. For example, Figure 7 shows the 1-D FFT menu in which each item is connected to a separate GX.

The GX language is designed for the skill level of a computing oriented geophysicist, not a computer scientist. This was an important design goal because we want geophysicists who best understand the applications to be the application developers. This also gives Oasis montaj users the ability to add their own functions and apply new techniques from their own research or from published literature.

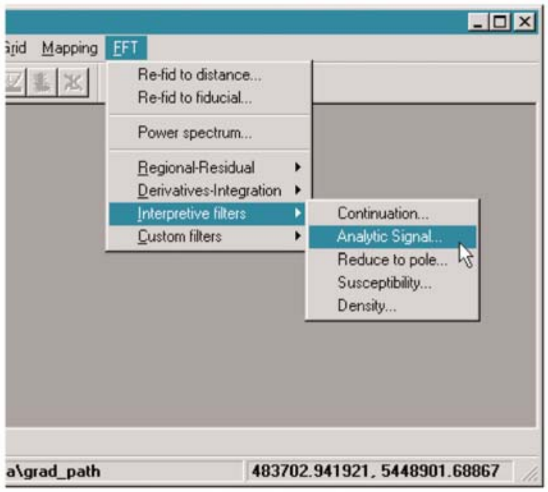

Figure 7: This shows the menu associated with the 1-D FFT GX suite. Each menu item is connected to a separate GX process.

GX is also used to connect Oasis montaj to other software systems and sources of data. This lets us, and others, integrate Oasis as a component in a more comprehensive computing environment. Thirdparty specialists, such as developers of geophysical modeling software, can also use GX to connect their systems directly to the Oasis system.

Dynamic Data and Process Links

The third key component in the design of Oasis montaj is the concept of links. We provide two kinds of links – dynamic data links and dynamic process links (also called "Makers").

Traditionally, when an interpreter sees an anomaly on an image, she is not able to easily see the original data that produced that anomaly. Nor is it readily apparent what process was involved in processing the data from it's original form to create the observed anomaly. Dynamic data links and dynamic process links address these problems.

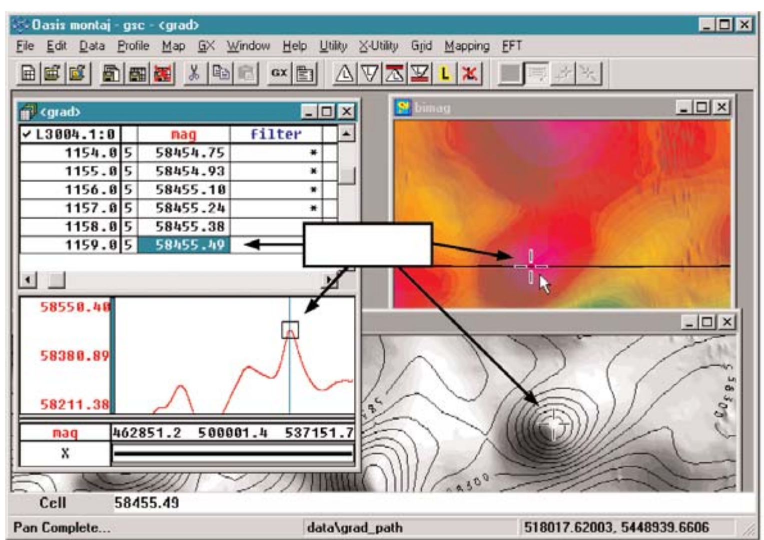

Figure 8: A data link allows an interpreter to select an anomaly on an image and immediately see the original data in the spreadsheet view and on other maps

Dynamic data links are graphical connections that link what a users sees on a map to the same location on other displayed maps and to the original data in spreadsheet and profile views. The interpreter who selects an anomaly on an image can immediately see the original data in the database and in profile form as shown in Figure 8. Link performance is critical and can be a concern for very large databases. However, the design of the Oasis database and the ability to imbed link information inside visible map entities provides a response that is almost instantaneous.

GEOSOFT

Technical Paper

Dynamic process links provide access to the process and process parameters that were used to create something. When you select a channel in a database, clicking on the right mouse button will display the “maker”, which will activate the GX process that created the channel together with the control parameters. The user can change the parameters and re-make the channel.

Figure 9 shows the “Shaded Color...” maker that is displayed when an image on a map is selected. When the “Shaded Color...” item is selected, the GX that is run and the GX dialogue appears. The user can change the parameters and the image will change accordingly.

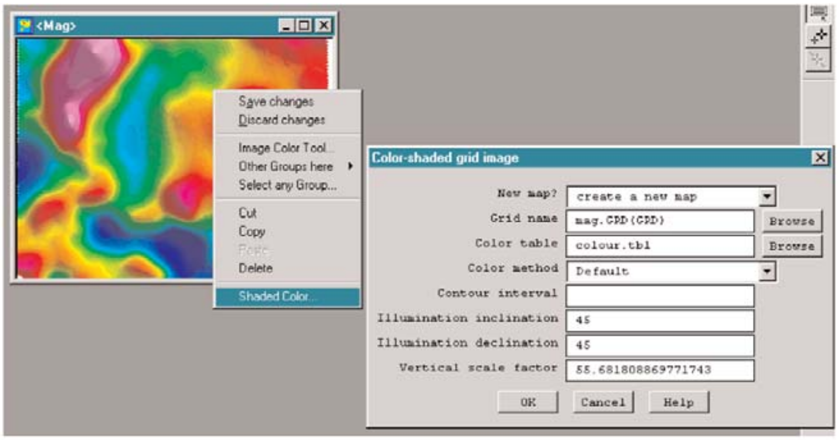

Figure 9: A Process Link is activated when an object on a map is selected. The object “maker” is added to the bottom of the properties menu, and selecting the maker will run the GX that created the object.

Summary

Oasis montaj represents a new generation of commercial software that is designed to address the special requirements of airborne data processing. It specifically:

- Addresses the requirement for large and efficient database storage of airborne data.
- Provides a graphical user interface that lets uses work with their data in spreadsheet, profile and map view.
- Provides a way to extend the system and add functionality so that new techniques can be added both by Geosoft, end users and third parties.
- Delivers both dynamic data links and dynamic process links so that users can relate data and the processes that have been applied to the data.

Oasis montaj provides a geophysicists with a way of maintaining contact with their data throughout the geophysical data life cycle – from original data storage to processing to extraction of meaningful geologic information and interpretation. As for the future, we at Geosoft are working to improve this system in the following areas:

1. Adding more connections (import, export and direct connection) to data from other applications, Geographic Information Systems and from corporate databases.
2. Adding new functions to meet the demands of our science as it evolves, both by developing new capabilities ourselves and by working with third party developers.
3. Improving techniques for educating potential users on how to effectively use the ever expanding capabilities of the system.
4. Addressing the need to work with ever larger databases and allow the sharing information among the members of project teams.

We see the continuing development of advanced software systems like Oasis montaj as an important part improving our ability to extract useful information from our data.