EVALUATION OF THE GLORIA CD-ROM DATA SYSTEM FOR THE GULF OF MEXICO PRODUCED BY THE U.S. GEOLOGICAL SURVEY AND THE NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

Stephen J. d'Apollonia

DATA QUEST
16 Major Street
Dartmouth, Nova Scotia B2X 1A4

Prepared for:

Ron Macnab
Atlantic Geoscience Centre
Geological Survey of Canada
Bedford Institute of Oceanography
Dartmouth, Nova Scotia

G.S.C. OPEN FILE REPORT NO. 2199

July 1989

ABSTRACT

Evaluation of a preliminary copy of GLORIA side scan sonar imagery for the Gulf of Mexico stored on CD-ROM (Compact Disc Read-Only-Memory), and which was produced jointly by the U.S. Geological Survey, NOAA and NASA, has shown that CD-ROM can be a highly suitable media for distribution of large scientific data sets in digital format. Data can be retrieved, manipulated and displayed using personal computer systems equipped with a CD-ROM drive. The GLORIA CD-ROM data system for the Gulf of Mexico provides a user-friendly capability for the retrieval and basic image processing of a compressed 900 megabyte data set of side scan sonar imagery on a single CD-ROM disc. All the sonar images on the CD-ROM data have been processed for geometric and radiometric corrections, and therefore applications requiring reprocessing of uncorrected GLORIA image data is not feasible using this data.

INTRODUCTION

This report summarizes a preliminary assessment that was carried out on an evaluation copy of processed GLORIA data from the Gulf of Mexico stored on CD-ROM. The disc was produced by the USGS/NOAA Joint Office for Mapping and Research in the Exclusive Economic Zone. The CD-ROM electronically reproduces the equivalent of about 900 megabytes of GLORIA side scan sonar imagery previously published as an atlas by the U.S. Geological Survey (USGS, 1987).

Because CD-ROM (Compact Disk Read Only Memory) is still in its infancy as an accepted medium for the distribution of scientific data sets, and because its potential has not been widely considered, this report is in two parts. Part 1 is an overview of the basic technology of optical storage, specifically the compact disc, and compares this medium to other forms of information storage. Part 2 summarizes an assessment of the usage and potential applications of the Gulf of Mexico GLORIA CD-ROM data for scientific analysis and data distribution.

PART 1 - CD-ROM - THE NATURE OF THE MEDIUM

Data Storage Technologies

Encoded optical storage has been used commercially since the late 1920's for recording motion picture soundtracks. This was followed by the development of magnetic recording which subsequently came into widespread use in the 1940's and 1950's. At first, magnetic recording was used to record analog information such as music, voice, and television signals. One characteristic of analog recordings is that the fidelity of the source material deteriorates to some degree with each generation of the original recording. This is in contrast to digital recording systems, such as used in computers, which do not suffer from the degradation that can effect analog information. The reason for this is that digital copies of the original recording are not an analogy or approximation but rather an exact bit-for-bit replication of the original data.

Magnetic tape is still a widely used standard for magnetic storage of digital information. However, the sequential file structure intrinsic to magnetic tape data systems results in long data access times. Therefore, in computers, magnetic tape is primarily used for archiving or transporting data.

The problem of long access time was greatly improved with the development of magnetic disks. These permit random mass data access to

fall to a fraction of a second. Magnetic disk storage systems include fixed hard disk drives such as the Winchester drive, removable hard disk systems such as the Bernoulli Box, and flexible disk systems commonly called floppy disk drives. Storage capacities of magnetic disk storage systems for personal computers range from 360 kilobytes to several hundred megabytes. They are ideally suited for on-line data storage, backup, archiving, and program distribution, and are common on most personal computer systems today. Since magnetic data systems have both read and write capability, data can be easily modified or erased. Disadvantages of magnetic disk systems used for data distribution are: 1) the inherent volatility of the data when subjected to magnetic fields; 2) the unprotected nature of the disk surface which can be easily damaged by mishandling; and 3) disk failures from head crashes that can result in the loss of all or a portion of data on a disk.

The development of optical disc technology became commercially successful with the product known as CD Audio. It offered superior audio characteristics, convenience, and high durability. CD Audio essentially carries an exact digital replication of the master recording. CD-ROM uses the same technology as CD Audio, the only difference being additional error correction to meet computer standards, a direct digital output, and certain mechanisms in the CD player to allow a computer interface.

Although other optical storage technologies exist, such as WORM (Write-Once-Read-Many times) drives and erasable optical disks, the CD-ROM has become a "medium of choice" for mass distribution of read-only information. There are two important reasons for this: 1) CD-ROM is

the most currently standardized form of optical disc media for computer use; and 2), it is the least expensive for mass production, because it takes advantage of the manufacturing technologies of CD Audio. The same plants that manufacture CD audio discs also produce CD-ROM discs. A single CD-ROM disc can store a nominal 540 megabytes of user data on a 12 cm disc. This is equivalent to 1,300 5.25-inch double density floppy diskettes or 200,000 printed pages of text. CD-ROM is ideally suited for distributing multiple copies of large data sets. Although information can be extracted and manipulated by users, the data on the CD-ROM disc cannot be overwritten or altered in any practical way. Table 1 summarizes data transfer and seek rates for CD-ROM and hard disk systems.

TABLE 1 - Seek and Data Transfer Rates for CD-ROM and Magnetic Hard Disk Drives

	CD-ROM	Low-Performance Hard Disk	High-Performance Hard Disk
CAPACITY	540 Mbytes	10 Mbytes	456 Mbytes
AVERAGE SEEK	500 ms	100 ms	28 ms
MAXIMUM SEEK	> 1 sec	200 ms	50 ms
ROTATIONAL SPEED	200-500 rpm	3600 rpm	3600 rpm
TRANSFER RATE (SEQUENTIAL)	150 Kbytes/s	96 Kbytes/s	312 Kbytes/s

(Source: Buddine, 1987)

CD-ROM As A Distribution Medium For Digital Data

Organizations which produce, use and distribute large volumes of information must consider a number of key questions in deciding if CD-ROM is an appropriate distribution medium. These questions include the following:

- What is the application, who is expected to use it, and how will the user benefit from how the data is presently being distributed?

- CD-ROM is a read-only, mass produced medium. What advantages and disadvantages does this present to the number of potential users and the level of data currency (updates) required.

- What will be the computer/CD-ROM system on which the application will run? What standards will be imposed on the delivery system?

- What is the cost/benefit of using a CD-ROM based application in terms of additional data value and cost of additional equipment?

- Who will be responsible for developing the application within the organization? How long is the project expected to take? What resources will be required? Who will support the system, provide documentation, training and maintenance, updates?

- Can the appropriate retrieval software be purchased or licensed, or will it have to be developed in-house?

- Will the project be carried out in-house or contracted out to a service bureau? What portion of the work can be effectively completed in-house?

The answers to these questions are critical in determining if the effort and cost of producing a CD-ROM application can be justified.

CD-ROM Data Fundamentals

Apart from its relative "newness" and lack of widespread availability and usage, CD-ROM has two important technical disadvantages over magnetic disk storage. First, in order to maximize the storage capacity of CD-ROM discs a CLV (constant linear velocity) recording format is used. This means that the disc slows down as it reads from the inside track to the outside track, maintaining a constant speed of 1.2 m per second under the read head over the entire disc. This is in contrast to magnetic disk systems which use a CAV (constant angular velocity) recording format, where the disk rotates at a constant speed. On magnetic disks (CAV), information density decreases from inner to outer tracks. CLV recording increases the data storage capacity on CD-ROM discs, but also increases data access time since the CD-ROM disc must change speeds as it moves to different positions over the disc radius. This results in about an order of magnitude increase in access time compared with magnetic disk systems. Because of this slower access time, data retrieval software designed

for magnetic storage systems give unsatisfactory results when used data stored on CD-ROM, and special designed retrieval software is required. The other disadvantage with CD-ROM is that it is limited to read-only applications. This information can include text data, computer programs, images and audio. For many applications the fact that CD-ROM is a read-only data system is not a disadvantage, but a desirable feature that ensures data quality, integrity, and copyright.

The CD-ROM disc uses the same data storage format as the CD audio disc. Data blocks are identified by minute, second and sector numbers. Data blocks are recorded along a spiral track from the inner to the outer radius of the disc. Each raw data block (sector) contains 304 bytes of housekeeping information and 2,048 (2K) of user data for a total of 2,352 bytes of information. Each data block is called a sector, and 75 sectors are stored per second. The nominal playing time of a CD audio disk is 60 minutes, and the data storage capacity of a CD-ROM disc is therefore 75 sectors/second x 60 seconds/minute x 60 minutes or 270,000 sectors. Thus the nominal data capacity of a CD-ROM is 540,000K.

CD-ROM is essentially a distribution medium rather than a storage medium. Single copies or very small runs (less than 50) have a high price per copy, thus it is only appropriate where a large number of copies are distributed in identical form. WORM drives, by contrast, have a high cost per disc and are therefore suitable as a single site storage medium, but inappropriate as a distribution system to a number of sites.

CD-ROM is considered a highly nonvolatile medium. The data cannot be accidentally erased or corrupted by exposure to a magnetic field. The data is not affected by the reading process, and the disc surface is highly durable and resistant to damage caused by dust or mishandling. The current life expectancy of optical discs is at least 10 years and probably greater.

The storage capacity of CD-ROM is a nominal 540 megabytes. For distribution of large volumes of information at a large number of sites, CD-ROM is currently the most inexpensive distribution medium of digital information. The delivery cost can be as low as $0.01/Mbyte compared to $1.00-$2.00/Mbyte for tape or floppy diskettes, and $5.00 or more for distribution by paper hard copy.

CD-ROM offers the potential of delivery of multimedia interactive applications to a microcomputer environment. This can include combinations of text-based applications, audio, graphics, images, and computer programs.

PART 2 - ASSESSMENT OF THE GULF OF MEXICO GLORIA CD-ROM DATA SYSTEM

Hardware Requirements and Installation

Minimum system requirements for the GLORIA CD-ROM are an IBM-compatible 8088, 80286 or 80386-based PC with a hard disk, one high density (1.2 Mbyte) floppy disk drive, at least 512K RAM, EGA graphics display device (640 x 350 resolution), and a CD-ROM drive with MS-DOS Extensions Version 2.0. The side scan sonar imagery data can be viewed in EGA graphics mode in either 16 colours or 4 grey levels. Although the 16 colours in EGA graphics mode are sufficient for producing false colour images, the 4 grey levels available in EGA mode are inadequate for viewing and analyzing the images in grey scale. A VGA graphics mode is also supported, providing higher resolution displays (640 x480) and 64 grey levels. This option requires installation of a high resolution graphic card (Designer VGA) available from Orchid Technology Inc.

The hardware system that was used for testing and evaluation of the GLORIA CD-ROM included the following components:

- IBM compatible 80286 microcomputer, 640K
- 30 M hard disk, 1.2 M floppy drive
- NEC multisync II monitor
- Orchid Designer VGA graphics card (Optional)
- Philips CM100 CD-ROM reader with MS-DOS Extension 2.0

Attempts to use the GLORIA CD-ROM disc on a NEC CDR-77 CD-ROM reader installed with a SCSI interface and NEC's MS-DOS extensions 2.0 were unsuccessful. The problem was subsequently traced to a software error in the device driver supplied with the NEC CD-ROM reader.

Gloria Data System Documentation

The GLORIA CD-ROM data system was supplied with a clearly written User Manual containing step-by-step instructions for installation and system use. Comprehensive additional documentation was provided in electronic form on floppy diskette including: names and addresses of contacts; GLORIA DATA Technical information; entire text of the Atlas of the U.S. EEZ, Gulf of Mexico; and User's Guide for the Planetary Data System IMDISP program. The documentation included a detailed explanation of all the file formats and data structures in each sub-directory of the CD-ROM.

The GLORIA CD-ROM data system was installed without difficulty using the written instructions that came with the system. The CD-ROM installation, display and retrieval software comes on a high density 1.2 Mbyte floppy disk. After the system is installed, the CD-ROM reader is recognized as another read-only drive. New CONFIG.SYS and AUTOEXEC.BAT files were written to handle the Philips CD-ROM reader and Orchid graphics card (Appendix A).

To access and display the GLORIA imagery data, a menu-driven user interface is provided. The CD-ROM contains three sets of processed digital sonar imagery data. These consist of 326 image files of raster data collected from the research vessel in any 6-hour time period, 16 image files created by electronically mosaicking the 6-hour files into composite 2-degree areas, and 2 image files representing the entire eastern and western halves of the Gulf of Mexico. To conserve space on the CD-ROM, the 6-hour image files are stored on the CD-ROM in a compressed format. A data decompression program called "PKARC" is provided to decompress the 6-hour files for viewing and analysis after downloading the file to the hard disk. The composite 2-degree areas are a duplication of the same areas in the published atlas for the Gulf of Mexico GLORIA data. The CD-ROM also contains bathymetric data digitized from NOAA's Bathymetric Map Series at scales of 1:250,000 and 1:1,000,000 that can be overlain on the sonar images. The bathymetric contours are in metres, with contour intervals of either 250m or 500m.

The installation procedure creates a sub-directory called GLORIA containing all the necessary program files to use the system. After installation, the menu system can be invoked at any time by going into the GLORIA sub-directory (cd\GLORIA) and entering: MENU.

Image Analysis Software

Two separate programs are provided with the GLORIA data system to display and analyze the image data. The first program (NOAAD) was

developed by NOAA, and enables the user to pick any 2-degree or alternate sub-area from an outline map of the Gulf of Mexico for subsequent display and analysis. The area is selected by moving the cursor to the desired area and hitting the return key. Images are initially displayed as a false colour image represented by the 16 colours of the EGA mode. Option A allows the user to select alternate default colour sets, interactively assign colour bands, perform colour band stretching, and perform 1-degree or window zooms to enlarge the displayed image. Option B allows the user to add bathymetric contours, or present the image as a 64 grey-level display (Orchid VGA Plot). All options using NOAAD were found to work as expected. Pixel resolution is given as 50m for the 2-degree images. When an image is displayed as a 64 grey-level plot (Option B), the representation on the monitor duplicates the black and white image of the same area in the published Gulf of Mexico Atlas, although at a somewhat reduced scale. The image analysis capabilities of the NOAAD software are limited to alternate colour band selection and colour stretching. The stretch option re-assigns the 16 available colours by regrouping less frequent DN (Digital Number) values for each pixel into a single colour and using the remaining colour bands to represent the more common DN values.

The other image analysis software provided with the GLORIA CD-ROM disc is called IMDISP, which was developed by NASA's Jet Propulsion Laboratory as an interactive image display utility for the IBM Personal Computer family and compatibles. IMDISP was initially developed for analysis and display of planetary image data on CD-ROM, but can also be used to display and manipulate images stored on floppy or hard disk systems.

IMDISP provides more comprehensive image analysis capabilities than is currently available with the NOAAD software. Image analysis can be performed on grey level images or on false-colour images. A few of the available command options are described below:

HELP	On-Line help message for available commands.
HISTOGRAM	Produces a plot showing the number of pixel values in the image for each DN value.
PALETTE	To modify the colour palette used to display the image.
PROFILE	To plot the DN values of pixels located along a line between any two selected points on the image.
STRETCH	To do a linear grey scale stretch of the image.

Any pixel on an image can be easily displayed and mapped in terms of its line (row) and sample (column) number. The DN value of each pixel is stored as 8 bit data (0-255), with black (0) representing the weakest acoustic return and white (255) representing the strongest acoustic return.

Discussion

Few difficulties were encountered with the software which was found to be easy to use and sufficiently documented for good basic comprehension. The 2-degree images were easily brought up and displayed from the index map. Retrieving the 6-Hour image files was somewhat more complicated as there was no index map from which files could be selected. John-Hughes Clarke, a post-doctoral researcher at the Atlantic Geoscience Centre with experience in processing GLORIA sonar imagery, concluded that the CD-ROM GLORIA 2-degree image data electronically duplicated the digitally processed sonograph images in the published Atlas for the Gulf of Mexico. From a scientific viewpoint, he was more interested in digitally unprocessed GLORIA data that would permit more flexible and comprehensive analyses to be performed.

The GLORIA CD-ROM data system does not provide a utility to generate hard copies of displayed images. Ms. Bonnie McGregor of the USGS was contacted for information regarding hard copy output. Subsequently, Mr. Millington Lockwood of the USGS recommended an inexpensive (less than $100) software package entitled " PIZAZZ + " which they have found works very well in producing high quality hard copy screen images in either colour or black and white.

From the point of view of the scientific researcher, the question that must be considered is the potential merits of digital data on CD-ROM. For example, in the case of side scan imagery, would it be possible to interpret and map surficial sediments and bedforms from digital images on a display monitor using image analysis techniques? Probably the answer is yes. But is CD-ROM an appropriate medium for this type of analysis? Probably the answer is no, if only a limited number of users (less than 50) are ever expected to use the data for that purpose. Basic image analysis techniques such as contrast stretching can be used to highlight and map subtle seabed features or areal deposits. But would doing so be more efficient or lead to more understanding and valid conclusions than by traditional methods of data posting and manual interpretation? The answer to this question will depend on the experience of researchers and the type of data available for interpretation. For many applications CD-ROM may be most cost-effective as a distribution media of final image/map/text products. The availability of unprocessed data products on CD-ROM may find additional uses for low cost data distribution to a much wider user community than is presently feasible.

The GLORIA CD-ROM delivers high quality processed imagery data together with retrieval and analysis software that allows the user to carry out basic image analysis functions. The product can be delivered in volume at low cost. The published Atlas that contains the same GLORIA Gulf of Mexico images is published in a high quality printing format, and contains additional seismic profile data.

ACKNOWLEDGEMENTS

The author wishes to thank Randy Currie and Paul Girouard of the Atlantic Geoscience Centre for their invaluable assistance and technical expertise in providing the computer system necessary to carry out this work. Sincere appreciation is also extended to Ron Macnab for his overall support and guidance, and John-Hughes Clark for helpful discussions. The work was financially supported by the Geological Survey of Canada.

REFERENCES

Buddine, L. and Young, E. 1987. The Brady Guide to CD-ROM. 476p. Prentice Hall Press, New York.

EEZ-SCAN 85 Scientific Staff. 1987. Atlas of the U.S. Exclusive Economic Zone, Gulf of Mexico. U.S. Geological Survey Miscellaneous Investigations Series I-1864-A,B 1987.

APPENDIX A - DOS FILE LISTINGS

The following DOS files were required with the particular computer hardware and CD-ROM reader used in this study:

FILE NAME: CONFIG.SYS

```
buffers=20
files=32
device = c:\cm153.msc /P:340 330
```

FILE NAME: AUTOEXEC.BAT

```
c:\mscdex /l:d /d:hsdvdmsc /m:8
path c\:;c:\util;c:\raf;c:\dos
set orchid = true
prompt $p$g
```