IN-43
198090
62 P

Publication of Science Data on CD-ROM: A Guide and Example

Gary Angelici and J. W. Skiles

NASA

National Aeronautics and
Space Administration

Publication of Science Data on CD-ROM: A Guide and Example

Gary Angelici

Sterling Software, Inc.
1121 San Antonio Road
Palo Alto, CA 94303

J. W. Skiles

Johnson Controls World Services
7315 N. Atlantic Ave.
Cape Canaveral, FL 32920

NASA

National Aeronautics and
Space Administration

Ames Research Center
Moffett Field, California 94035-1000

TABLE OF CONTENTS

Section 0   SCOPE

0.1   Scope

This document is intended to be a guide for the publishing of science data on CD-ROM (Compact Disc-Read Only Memory), offering lessons as learned from the publication of data for a NASA ecosystem science project. It covers the entire procedure, from the determination of publication feasibility to the distribution of discs to the customer. While each publication situation is unique, this document will offer a sequence of steps that should be performed and questions that should be answered in the process of science data publication.

0.2   Revision Level

This is the original version and is given a revision level of 1.0. Future versions will be numbered according to the extent of revision.

Section 1   INTRODUCTION

1.1   Introduction

Thousands of CD-ROMs have been published by various commercial and governmental organizations as a means of disseminating scientific data and information. These discs have been largely praised for their ability to get large volumes of the most important data out to practitioners, researchers and educators in an economical and easy-to-use fashion. For little or no money, data which would otherwise be in a distant difficult-to-access magnetic tape archive can be slipped into a device that can be connected to various computers and displayed and processed. With CD-ROM readers becoming almost standard equipment on many computers, the popularity of CD-ROMs is certain to rise.

Few documents have been written to describe the entire procedure of producing a CD-ROM, and they rarely include data and documentation preparation, which are often the most time-consuming, expensive and important aspects of production. This document seeks to partially fill this void with detailed information about and examples for the several activities necessary in the production of CD-ROMs. The many decisions that have to be made, including the "obvious" and the difficult, options that are available, and answers to many questions will be enumerated. If one lesson is to be learned from this document, it is that there are many details which must be considered in the publication of data on CD-ROM and that it can be a labor-intensive activity.

This document describes, basically in chronological order, each step that should be accomplished in the production of CD-ROMs, from feasibility determination to disc distribution. Important aspects of each step are described along the way, including standards, procurement, costs, and other variables. While this document attempts to be complete, each CD-ROM will be different and different techniques will undoubtedly have to be applied. For example, different procurement situations and types of data may require more or fewer tasks to be performed. Different methods of premastering data (which may be performed on different platforms with different software) will require unique solutions. Because of these differences and the fact that the author does not have experience in all types of data publication, this document cannot be considered an all-inclusive dissertation on the subject of data publication. Likewise, because large volumes of CD-ROMs have not been produced by the authors, this document is almost by definition a preliminary document. More needs to be learned and, with allocation of time, incorporated into improved versions of this document.

1

In addition, the sheer volume of decisions that had to be made prevents complete documentation of all aspects of CD-ROM publication. The intent of the document is to give a broad overview and highlight the important issues of the basic steps involved. One issue which will not be covered, for example, is the optimum skills profile of the staff publishing CD-ROMs, although an idea of that mix can be discerned from the text.

It is recommended that the entire document be read before embarking on the road to publishing data on CD-ROM. No only will it help the reader evaluate whether it is desirable to create a CD-ROM, but it may identify some purchasing or staffing decisions that may have to be made in the short term. Once the procedure is understood, this document can be referred to for details at each step of the way in generating a CD-ROM.

The coordinator for the production of a CD-ROM for an earth science project, the Oregon Transect Ecosystem Research (OTTER) project is the chief author of this document. OTTER collected hundreds of spectral and photographic images taken during satellite and aircraft overflights and hundreds of measurements (stored in the form of tables) obtained from a wide variety of ground-based instruments. Documentation to describe the data was also written by investigators. Much of these scientific data and documents were prepared and published on CD-ROM during 1992. As of this writing, the data for volumes two through five have been premastered and are being reviewed by scientists and CD-ROM experts. The text in this document emanates from the experiences of producing the OTTER CD-ROM. Lessons learned in the production of the OTTER CD-ROM are offered throughout this document to clarify issues, which are intended to aid others in the production of their CD-ROMs. These experiences are typically included as the last paragraph of applicable sections.

The persons responsible for the publication of data may or may not be the persons responsible for data collection and analysis. Intimate familiarity with the data and its uses is an advantage for the persons who will publish the data, particularly when it comes to data documentation and verification. The OTTER project data was published by persons with little knowledge of the collection or analysis techniques. This document will be of equal use to both groups of data publishers.

Each Technical Note should have a Scope statement (Section 0.1) and an Introductory statement (Section 1.1). The management summary (below) is optional, and is usually only used in cases of a Technical Note as documentation for a task's organization and specifics on work done for that task.

1.2   Management Summary

The following flow chart depicts the steps necessary in producing a scientific CD-ROM, with boxes at the same horizontal level denoting that the tasks can be performed simultaneously. This is the recommended order of step execution, but, in reality, circumstances will modify the order somewhat. For example, just because the entire staff is not familiar with CD-ROM technology doesn't mean that the task of data selection cannot begin.

```
                    ┌─────────────────────────────┐
                    │  Data Publication Feasibility │
                    └─────────────────────────────┘
                                   │
          ┌────────────────────────┼────────────────────────┐
          ▼                        ▼                        ▼
  ┌───────────────┐        ┌───────────────┐        ┌───────────────┐
  │    CD-ROM     │        │    Market     │        │    Design     │
  │ Familiarization│       │   Analysis    │        │  Evaluation   │
  └───────────────┘        └───────────────┘        └───────────────┘
          │                        │                        │
          └────────────────────────┼────────────────────────┘
                                   ▼
                          ┌───────────────┐
                          │ Data Selection │
                          └───────────────┘
                                   │
                                   ▼
                          ┌───────────────┐
                          │    CD-ROM     │
                          │    Design     │
                          └───────────────┘
                                   │
          ┌────────────────────────┴────────────────────────┐
          ▼                                                  ▼
  ┌───────────────┐                                  ┌──────────────────┐
  │     File      │                                  │  Disc and Insert │
  │  Preparation  │                                  │  Artwork Design  │
  └───────────────┘                                  │  and Production  │
          │                                          └──────────────────┘
          ▼                                                  │
  ┌───────────────┐                                          │
  │     File      │                                          │
  │  Verification │                                          │
  └───────────────┘                                          │
          │                                                  │
          ▼                                                  │
  ┌───────────────┐                                          │
  │  Proof Disc   │                                          │
  │ Premastering/ │                                          │
  │   Mastering   │                                          │
  └───────────────┘                                          │
          │                                                  │
          ▼                                                  │
  ┌───────────────┐                                          │
  │  Proof Disc   │                                          │
  │    Review     │                                          │
  └───────────────┘                                          │
          └────────────────────────┬────────────────────────┘
                                   ▼
                          ┌───────────────┐
                          │  Final Disc   │
                          │  Mastering    │
                          └───────────────┘
                                   │
                                   ▼
                          ┌───────────────┐
                          │  Final Disc   │
                          │ Distribution  │
                          └───────────────┘
```
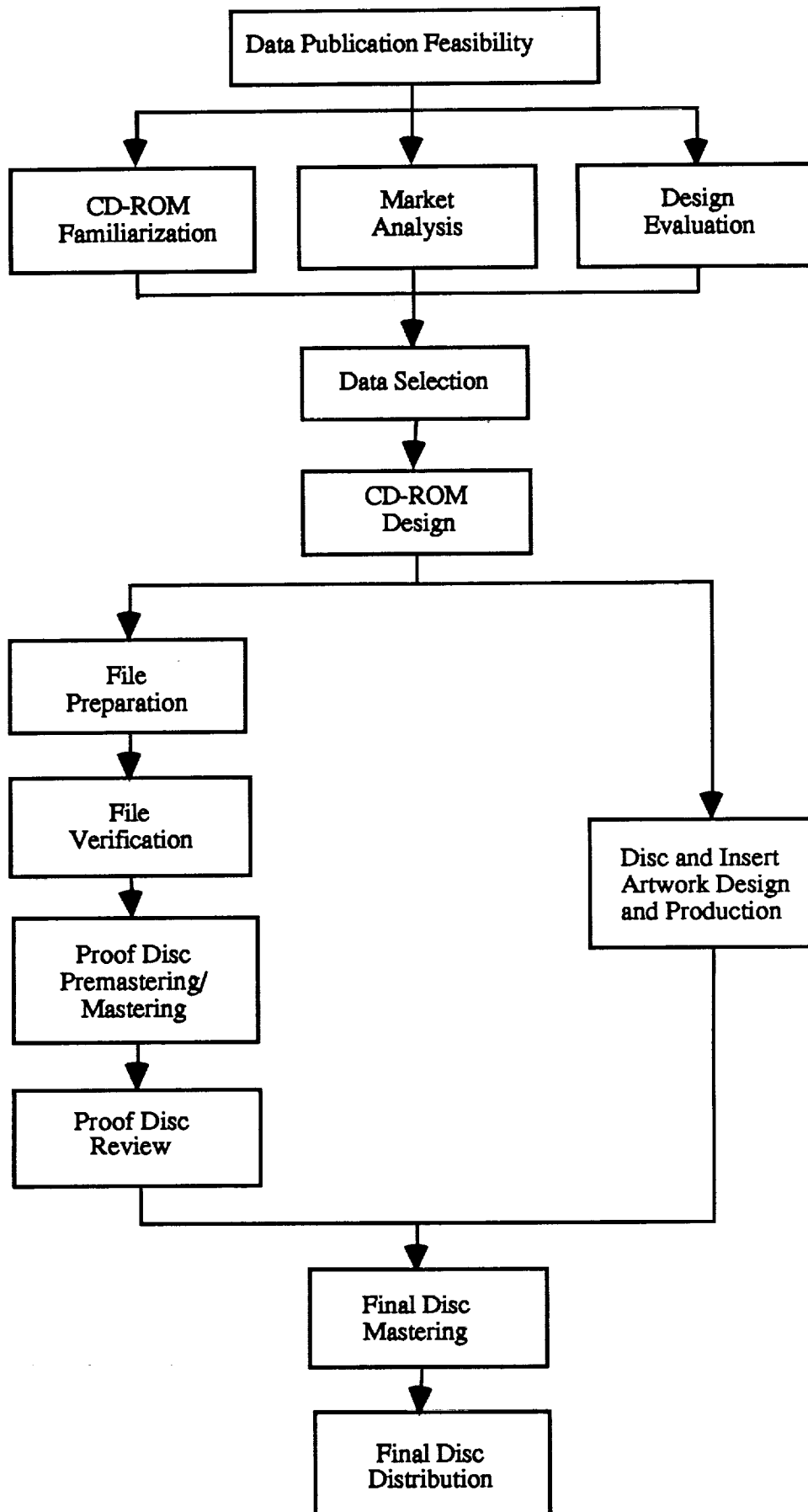
Section 2   **DATA PUBLICATION FEASIBILITY**

2.1   Introduction

It is important to be absolutely sure that CD-ROMs should be produced prior to embarking on that path. It is typically a very time (and money) consuming task which requires a major commitment from several parties. This is not the time to be overcome by the popularity or high technology nature of these shiny little discs. If an expenditure of thousands of dollars is to be made in producing the discs, it is essential that the groundwork be performed in its entirety. This section covers those data publication feasibility tasks.

The first paragraph mark on the title page is set at (line -12, after 82). The template shows the format of the information on the title page. The technical note number is derived as follows: TN for technical note; the calendar year the document is published; 8XXX, with XXX being the three digit task number (example - task 8 is 008); YYY being the subtask number, if any (if none, the YYY should be 000); and Z being the sequential number assigned to this particular technical note by project support. The title is the title. The originator is the person or people who wrote the technical note. The date is the actual date formatting and proofing are completed, and the tech note is ready for publication. This date need not be a specific day, but at lease a month and year. The footer for the title page section is center justified, and contains Sterling's address and phone number.

2.2   Demand

The most important determinant of data publication feasibility is the demand for the data. There are two potential sources of this demand: a market which will pay for the data, or a request for publication from a funding official. The author is not familiar with market analysis and will defer to business experts for the determination of a market which will pay for the data. In the scientific environment, it is typical for a funding official to request that the data be published as a means of recording the collection of the data in a cohesive and easy-to-use package for distribution to a larger scientific community. The official feels that the data is important enough to the general scientific community to spend funds for the preparation of data and mastering of the discs. The discs are typically distributed to the scientific community free-of-charge and through a government office for sale to the educational and commercial markets.

The OTTER project was funded to create the OTTER CD-ROMs based on the strength of the market as perceived by NASA Headquarters managers and based on the popularity and usefulness of discs produced by similar projects in the past.
The section headings are the table of contents, and should be preceded by a "C" marker as shown in this document. Normally, the "C" markers would be hidden text, but for the sake of example I have printed them here. Spacing between first level subsections should be larger than spacing between first and second, or second and third level subheadings.

2.3   Publication Media

The volume of the data that is to be published and the platform on which the data will be used largely determines the medium on which the data should be distributed. If only a few megabytes of data are involved and the potential users are expected to use Macintosh or PC platforms, then floppy diskettes may be the most appropriate medium for distribution. 3 1/2" floppy diskette drives are available on most personal computers, which have the power to display and process

large volumes of data. Data sets consisting entirely of tabular data and accompanying documentation are good candidates for the floppy diskette distribution approach.

If a large number of files is involved or if the files are large, a larger format medium is typically required. Larger format media, such as 9 track, 1/2", and 8mm magnetic tape (among others), may be appropriate for distribution if all targeted users have platforms with that common data reading drive. Since it is becoming a standard (or close to a standard) on a variety of platforms, CD-ROM format is the more reasonable choice. Although CD-ROM allows the publication of up to 650 megabytes of data on one disc, it is entirely acceptable to put much less data on CD-ROM. And if the target market is limited to a few users, it is possible to publish data on Write Once (WO) CD-ROMs, lowering the cost of publication greatly.

Because approximately three gigabytes of OTTER data were scheduled to be published, multiple CD-ROMs were seen as the most effective publication media.

2.4   Cost

The cost involved is always an important determinant in the feasibility of doing or producing anything. The best way to estimate the cost of performing any large effort is to break down the work into separate tasks and estimate the manpower and other resources necessary to complete each task. To aid in the estimation of the costs of producing a CD-ROM, the remainder of this document breaks down the work involved into specific tasks. As an example, estimates of the time and costs involved in the first OTTER CD-ROM is broken down by type of expense, by type of data file, and by preparation stage in section 15. Of course, each situation is unique and will require individual evaluation of time and cost. To be better able to estimate the cost involved in producing a CD-ROM, the entire document should first be read.

Section 3   CD-ROM FAMILIARIZATION

If the staff that is to produce the CD-ROM is unfamiliar with this relatively new technology, then a period of familiarization with CD-ROM technology relative to the target data is the first step. And as with all new technology, it continues to evolve and change.

The latest information regarding CD-ROM publishing should be read from up-to-date magazines, textbooks, and other sources. Experts in the area should be contacted, and sample discs from other data publishers who produced similar discs should be contacted for advice and documentation. If it is explained that it will be used as a guide in creating a different CD-ROM, a complimentary copy of discs from these other publishers may be requested. A CD-ROM drive should be purchased or access to an existing drive should be obtained to test various sample CD-ROMs.

Section 4   MARKET ANALYSIS

Before any product can be introduced into the market, a thorough market analysis should be performed. It must be assured that the product will be created so that it will be purchased (or used, in the case of "free" scientific data discs). The product must be tailored to satisfy the needs of and be easily used by the specific market that is being targeted. To do this, the characteristics of the prospective users must be understood.

To identify the characteristics of CD-ROM users, it must be known how the users will use the data and software. This knowledge will drive how the data and software will be configured on the CD-ROM. Typical questions about prospective users that should be asked include:

5

1) How broad is the customer base? Will just the persons who created the data be using the CD-ROM, or will a broader community be using the disc? If the latter, in what major fields of research or industries will this product be used? This will affect the language used on the text files of the disc, the amount and nature of the documentation, and other characteristics of the disc.

2) What hardware and software are used by the users? Do they typically use one piece of specialized hardware and one software package for these data or can they have a wide range of hardware and software to do their analyses? Do they typically own or have access to CD-ROM drives? For the OTTER data, remote sensing scientists in the ecosystem research field increasingly use IBM PC's (or compatibles) and Macintoshes for the display of data and Sun workstations and VAX minicomputers for image and statistical processing. Spreadsheet software on the personal computers is also used heavily by investigators to process tabular data.

3) What format will be most useful to the users? Related to item 2, the software drives which formats are chosen. If specialized software is used, it is best to provide the data in the format required by the software. If it appears that the software can read data in a more generic format, it is best to provide the data in the more generic format.

4) What is the level of sophistication of the users? Will they know how to use complex software if it were placed on the CD-ROM? Do they need additional information, such as calibration data on the disc as well? On the OTTER CD-ROM, ancillary data on about each flight line will help investigators interpret the image values.

5) For what exact purpose will (or could) the users be using the data and software? For example, will they need latitude and longitude coordinates of the corners of the image for registration purposes? On the OTTER CD-ROM, for example, ancillary atmospheric correction data can be used to correct the image data on the disc.

Keep in mind that markets (and their unique characteristics) exist that cannot be foreseen when data is first published. A market that cannot be easily characterized but will undoubtedly exist in the future is the education market. Schools from elementary school level to graduate college level may someday be viewing discs that may have been created for a very different market.

Section 5   DESIGN EVALUATION

The structure of the CD-ROM and the data files to be put on the disc depends largely upon the market for the disc and the standards, if any, that this market expects. Some markets are very precise in the data formats and structures that should be applied to data being published, such as astronomy market which uses the data structures of the Planetary Data System for the publication of deep space data. But most target markets are more diffuse and have not settled on a standard or are not monopolized by any one group. Such a market is the earth science market. Since no standards have been established either by agreement or general popularity, the structures of a CD-ROM for the earth science market, at this point in time, are at the discretion of the designer.

In this section, we review some actual standards of CD-ROM publishing of scientific data (ISO 9660), make some recommendations for CD-ROM design, and give examples of differing designs, including the OTTER choices. Because of the newness of the technology and the entry of new institutions publishing data, the best designs are always changing and new and better designs are always emerging. In the article written by the Jet Propulsion Laboratory's Data Distribution Laboratory (1993), details on the formatting of data for publication can be found. The CD-ROM design "market" should be thoroughly reviewed by talking with experts in the

field before adopting any design. Up-to-date periodicals, such as those listed in the appendices should be read for the latest in technology and data publishing standards and designs.

5.1   Volume and Directory Structure

One well-accepted standard for the structure of volumes and directories of CD-ROMs is what is known as the "ISO 9660" standard. This standard was specified by the International Organization for Standardization (ISO) and has the formal name (which should be placed on all discs which conform to the standard) of:

Information processing -- Volume and file structure of CD-ROM for information interchange, 1987, ISO/DIS document number 9660, International Organization for Standardization, 1 Rue de Varembe, Case Postale 56, CH1121 Geneva 20, Switzerland.

This standard, common in the publication of scientific data, permits data to be read from discs mounted on CD-ROM players which are capable of reading ISO 9660 standard discs regardless of computing platform. To achieve this level of portability, this dictates that all text information will be stored in a common format, which in this case is ASCII format. It also states that file names must conform to the DOS file naming convention of eight characters followed by a three character extension.

If the ISO 9660 standard is not necessary (multiple platforms will not have to be accommodated), one only needs to consider the rules of the target system. But since science has not established a "standard" computing platform, the use of the ISO 9660 standard with all of its restrictions must typically be adopted.

Although the OTTER data was prepared and premastered on a Unix machine, all ISO 9660 specifications were observed, in order to make the discs more generally useful.

5.2   Computing Platforms

The choice of computing platforms on which the CD-ROM will operate is perhaps the most important decision to be made. It will affect everything from software selection to disc testing. While this is largely driven by the platforms of choice in the target user market, the disc publisher can choose to add platforms which seem to compliment the data best or which seem to be an emerging standard, or the publisher can remove platforms which are felt to be less important. These decisions should be made carefully, as the addition of new platforms increases disc development costs and time lines and the removal of platforms can alienate important segments of the market. A workable balance should be struck between the two extremes by providing useful data to the majority of the users.

The platforms chosen as hosts for the OTTER discs were those most used by the earth science community: IBM PC's and compatibles, Macintosh, Unix systems, and VAX/VMS.

5.3   Data File Presentation

The data file presentation is the manner in which the data files are presented to the user. This can involve the naming of files, the format of the files, the techniques to describe the files, and other characteristics. There are three basic options: 1) one can structure the data in whatever fashion makes most sense to the designer or will be acceptable by the market or 2) one can adopt a developed data file presentation scheme, which may or may not be "standard", that has been applied previously in the publication of CD-ROMs, or 3) one can adopt the basics of a developed data file presentation scheme and modify it for the particular CD-ROM. All have been applied in the earth science arena with good results.

7

Presenting the data according to the designer's perception of the proper structure has very few rules and allows the most flexibility. One can tailor the CD-ROM to the data and to the needs of the user closely. This was done successfully in the FIFE Prototype CD-ROM, which included interface software which allowed PC-compatible users to easily access data organized according to a grid over the site. All characters of the file name, including the file name extension, were used to uniquely name each file within a directory. A disadvantage was that, although the disc was in ISO 9660 format, users on the Macintosh could not use the special interface and that there was no Macintosh display software on the disc.

The data file presentation scheme can be a complete system with rules for file format documentation, highly structured description files or headers with data dictionaries of fields to describe the data files, and tools to create its structures. An example of this is the Planetary Data System's (PDS) scheme of data presentation. These types of structures have the advantage of consistency across all data types and the support of large data system staffs which are improving the structure and writing software to serve the structure. Programs such as IMDISP, the public domain IBM PC-compatible display program, have been modified to automatically display an image when the accompanying label file (which describes the structure of the file) is selected. The disadvantages include the rather steep learning curve for the CD-ROM publisher to apply the structure to their data and for the user in understanding the language which may not be entirely from the domain of the data. For example, the language of the planetary community may not be easy for the earth science data publisher or user to understand. Another disadvantage is that the PDS standard dictates the use of the three character extension as file type identifiers, so these characters cannot be used for other information. It is recommended to, if available, use a reasonable data file presentation structure that fits the discipline and structure of the data. However, it would be better to do without a scheme than to have to modify the data files to the extent that they are difficult to read or understand.

If a data presentation structure is selected, required aspects must be decided, such as the fields to be put into the PDS label files. These, as always, should be selected with the data and user community in mind. It is advisable to include fields that are descriptive of the data from a scientific standpoint. Examples of important fields to include are location, date/time, altitude and pixel size (for imagery), and frequency of observation (for tabular data, such as meteorological data).

For the OTTER data, the third option was selected. The PDS data structure was adopted with some modifications. The PDS labeling feature was retained to take advantage of the easy display capability, while the data dictionary of PDS, written for planetary data applications, was modified to use earth science terminology where the planetary fields were not appropriate.

For each of the 1700 OTTER data files on the first CD-ROM, PDS label files were generated (most created by software).

5.4 Data Structure

Data structure issues deal with how the data are stored on the CD-ROMs. If binary data must be stored, the byte ordering may have to be considered. If text information is to be stored, the possibility of using well-known (but perhaps clumsy) standards must be evaluated.

5.4.1 Binary Data

An issue that must be addressed relates to the possibility of multiple-byte binary data storage modes and the effect on different platforms. This comes into play in particular with images. If each value of binary data exceeds one byte in length, the bytes are read in different order on

different computers. Whenever possible, it is recommended to avoid data files containing data values that consist of more than one byte or, if possible, to reduce multiple byte data to single byte data. But often data will lose its uniqueness if such a reduction is undertaken and the full two byte value must be preserved.

In placing the data on the CD-ROM, one can either have two versions of the data on the disc (one for each ordering) or have one ordering and record in the documentation the byte ordering along with a note that the bytes will have to be swapped on certain types of machines. Because the amount of data is probably large, it is advised to perform the latter option.

For the OTTER project, image data was kept in its 16 bit configuration to maximize information content. The documentation stated, given the platform, how the data were to be processed prior to use.

5.4.2   Text Data

Whenever possible, it is strongly recommended that data be stored in "standard" ASCII format. A major advantage of ASCII format is that it is portable between widely varying computer systems and software. It can be easily read and manipulated by nearly all word processing programs, including line editors, Unix "vi", VMS' "EDT" as well as personal computer programs such as Microsoft Word.

Most alphanumeric files, such as documentation, header, format, summary and other chiefly alphabetic files, are natural and obvious candidates for the ASCII format. There are also other types of mainly numeric data that can be stored in ASCII format. This includes data files that are in row-column tabular (or spreadsheet) format and may involve thousands of values.

It is possible to store text data in a common word processor format, such as Microsoft Word, if it is expected that most users will have this application. This makes it much easier to manipulate if the user will stay in that application. However, if the user wishes to use the same data with another application, such as a spreadsheet program, data transfer and loading becomes problematic. On the other hand, ASCII data is recognized worldwide and can be imported relatively easy into most software packages.

Even though most of the documentation and text files were generated on Macintosh computers in Microsoft Word format, the ASCII format was selected for CD-ROM storage.

5.5   Data File Format

The industry or scientific discipline which will use the disc may have its own formats for data storage and its own (perhaps specialized) software to process these data. Increasingly, however, the use of generic software tools, such as spreadsheet software on personal computers, is becoming more common in many industries and scientific disciplines. These tools are capable of using data files that are either in a "vanilla" format (as simple as possible) or in formats that are becoming accepted by a variety of groups. One such format that is becoming popular (as of this writing) is the Hierarchical Data Format (HDF) developed by the National Center for Supercomputing Applications (NCSA). It offers formats for binary and textual data that have been adopted by several prominent groups, including the EOSDIS (Earth Observing System Data and Information System). A thorough review of possible formats for the types of data that will be published should be performed prior to selecting a format.

5.5.1   Binary Data

Storing the data in its native format gives the user the peace of mind that the data are the original

data and have not undergone a (perhaps incorrect) re-formatting. In addition, software that has been written to read and process the data will still work with this unchanged data. At the same time, however, there are many users who may have difficulty understanding this (perhaps overly complicated) format and have no software to read the data. Even if they had the software, it may only run on a computer that the user does not have. This native format may not be able to be imported into popular display and image processing systems that are becoming "standard" in the industry.

Another option is to re-format the data into a common or "standard" format. Unfortunately, there are several standards in the market and none have been accepted in most domains. The market analysis should point to the most commonly used formats for the target user community. For image display, TIFF and GIF formats emerge as the most popular in the personal computer (and somewhat in the workstation) market. For image processing in the workstation market, files in the formats developed by image processing vendors such as ERDAS and ARC/INFO can usually be loaded easily.

Arguments in favor of the TIFF format as an image standard: It's a *de facto* standard, has lossless and lossy compression modes, and will handle 24 bit color or 8 bit greyscale images. Most image systems and many desktop publishing/graphics/whatever systems read and write TIFF files. The TIFF specifications are in the public domain, and implementations of readers/writers can be found on most of the large archive servers. 'tifflib' is one such piece of code, used widely by graphics programs.

Within each of these formats in the image community, there are also choices to be made relative to pixel ordering. Band Sequential (BSQ), BIL (Band Interleaved by Line) and PIL (Pixel Interleaved by Line) are pixel ordering schemes that each have their strengths and weaknesses. Once again, knowledge of the target market will identify which ordering is best for the CD-ROM.

A final option is to put the data on the disc in a "vanilla" form, which is byte data in a separate band per file format with no header. For image applications, each band of data is stored in a separate file, which completely avoids the issue of pixel ordering. This format has the advantages of easy understanding, easy reading by newly-written image processing programs, and easy import by most display and image processing programs. Disadvantages include the need to manage many more files on the disc and the inefficiency in processing separate files in software.

For the OTTER discs, the native format was selected for image data with complex formats and with hyperspectral characteristics. The cost of re-formatting the data exceeded the perceived benefit to the user community in terms of ease-of-use. For simpler files and for those image data types that required extraction of header information for the creation of ancillary files, the "vanilla" format was selected. Since considerable processing had to be performed to extract ancillary file information, the data was written back to disk in a more easy-to-use format.

5.5.2 Text Data

The choice of ASCII format as the data structure (see above) does not stipulate how the text and tabular data in the files will be formatted. For most complete portability of tabular ASCII files to spreadsheet and database programs on PCs and Macintoshes, two basic formatting rules are recommended:
1) End each field with a comma (not a tab or space)
2) Enclose each text field in double quotes

To improve readability without sacrificing portability, it is also recommended to start each field

in specific column numbers to improve readability.

In addition, different computer systems require different line ending protocols for all text files. Once again, the IBM personal computer and compatibles drive the format. For the files to be read by basic PCs (without special utilities), a carriage return and line feed must be placed at the end of each line. This can be accomplished by writing software or, if on the Unix system, using the "unix2dos" command. When complete, the file can be displayed legibly by most Macintosh word processing applications, by the DOS "type" command, and by the Unix "vi" word processor. Control M's (^M) appear at the end of each line on the Unix system.

The OTTER CD-ROMs followed all of these rules. Unix shell scripts were often written to generate text files with commas at the end of fields with fields beginning at the same column number. Also, software was written to place carriage return and line feed characters in positions 79 and 80 for documentation files.

5.6   Ancillary Files

Ancillary data files are those files which are necessary to understand the main data. These may include, for imagery, calibration and navigation data. For all types of data, header or format files may be necessary to adequately describe the data. These files are sometimes provided with the data and sometimes will have to be generated. Files for data presentation structures, such as PDS label files, were covered in the section on data presentation structures.

If there are none among the data publishing staff, experts knowledgeable in the data that will be published should be contacted for the ancillary files, software, and information that should be included with each data set. These experts are typically found where the data originated or can be found by contacting the data originators. Project scientists or those who provided the data should also be asked for the same inputs, because many times there are some special considerations for these particular data sets which do not apply to the data type as a whole. An example would be additional processing that a scientist performed on the data.

For the data on the first OTTER disc, there were calibration and navigation housekeeping data stored in the first bytes of each line of each band of data for most of the aircraft imagery. It was decided to copy the data to a separate ASCII file for each band of each image. Format files for these ancillary files were also generated to describe the format of these housekeeping data files. In addition, a file summarizing general, calibration, and locational data was deemed necessary for each scene. This configuration of files is similar to the group of files produced for the FIFE Prototype CD-ROM. While this configuration should be determined at the data selection step when discussing which imagery should be on the disc, certain types of ancillary files are necessary for selected types of data.

5.7   Data Compression

If several large format images are due to be published, great economies can be obtained by compressing the images. At this stage of the procedure, it is enough to be aware of the data compression algorithms available. When selecting data for inclusion on the CD-ROM, the rate of data compression should be known, in order to be able to estimate the volume of data for publishing. Knowledge of the target user market will determine whether compressed data on the disc will be accepted and whether it will be necessary to make de-compression as transparent as possible for the user.

At this time, lossless compression algorithms (Newcomer, 1992) offers the best solution for the compression of image data. Since this is a field which is rapidly changing, other such algorithms should also be investigated.

11

After investigating the use of a lossless compression routine used by the FIFE data publishers, it was decided not to compress the OTTER data for volumes 2 through 5. The output format of the files was incompatible with the format of the files on volume 1, and the procedure to uncompress the files was judged too cumbersome for the user.

5.8   Software

Helping the CD-ROM user to access, understand and use the data should be a goal of every person in charge of publishing science data. Carefully-selected software published on the CD-ROM itself is an important means of accomplishing this goal. The nature of the data and the market analysis of the target user community should dictate the type of software that should be on the disc.

A user interface to allow the access of all data for a particular time or location could help the user to use the data more quickly. The FIFE CD-ROM Prototype had PC user interface software, written by the CD-ROM development staff, which allowed access to all types of data for each grid cell of the FIFE site. This has been praised by investigators as a straightforward method of assembling relevant data.

If the data are mostly imagery, then software to display the imagery is a good way to help the user become familiar with the data. Software which performs basic processing of images (perhaps as part of the display program) would also be helpful for most users. Instrument data would benefit from calibration software on the disc, if the data have not already been calibrated. For some data types, the imagery in its raw form cannot be displayed using any software. With software published on the CD-ROM, the user, allowed to select different options such as band number or other features, can convert the imagery to a format which can be displayed. Software to carry the processing of the imagery to the next step, such as the atmospheric correction of the data on the disc, would also be beneficial.

While scientists are increasingly using spreadsheet software on their personal computers, tabular data could also be manipulated by software published on the disc. Specialized and complicated software to perform popular calculations would probably be appreciated by many users of the disc. Once again, the market analysis should uncover any useful software or procedures that users would appreciate.

Most high performance software was developed by private companies, is copyrighted, and therefore cannot be placed on CD-ROMs for distribution. That would be a quick way to be the object of a lawsuit. There is, however, a large body of software written by the government and others which is public domain and can be published on CD-ROM. It is usually possible to find some public domain software which performs the functions that are necessary in processing the data on the disc. This software is typically less powerful and less refined than commercial software, but will do the functions required if selected carefully and tested fully. These software programs can be identified by contacting major CD-ROM publishing facilities, such as the JPL Data Distribution Laboratory, that deal with the same kind of data.

As a word of warning, there are some platforms (notably Macintosh) that require that the program executable be premastered on the system of the executable. For example, to include the actual executable of any Macintosh program (commercial or public domain) such as Image4pds on a CD-ROM, the data must be premastered using premastering software on a Macintosh. As an alternative, a BINHEX version of the executable can be published on the CD-ROM with instructions to the user to obtain the shareware package, Stuffit, which can be used to generate the executable. The BINHEX version of public domain software can be obtained from major CD-ROM publishing facilities.

12

On the first OTTER CD-ROM, IMDISP software for the PC and Image4pds software for the Macintosh is stored on the disc for the display of images. On one of the other four volumes, software entitled SYNTHESIZE allows users to extract polarizations of data from the highly complex radar data stored in compressed format.

5.9   Conclusion

Rather than charge a committee with the task, it is best to allow one person to make decisions about the standards to which the disc will adhere and the design the disc will have. This person should know about all standards available to adopt, know the capabilities (both cyber and cognitive) of the target users, know the data, and be aware of budget limitations. Once a set of standards is selected, they should be presented (with rationale for selection) to another group of knowledgeable people. The standards should be reviewed for plausibility and usefulness by persons who are intimate with the data and familiar with the user community. A set of standards that will provide effective access to the data without encumbering the user with cryptic text in the files should be the result of the process.

In the OTTER experience, the author was the standard and information gatherer who made the decisions regarding the structure of the discs and files. The author consulted experts at various data publication sites for advice in various issues, many of which were discussed above. Except for the ISO 9660 standard, the author found that there are few expectations in the community about these design considerations. This may be attributable to the relative newness of the technology, and more expectations may surface with when the number of discs on the market increases.

<div align="center">

Section 6   **DATA SELECTION**

</div>

The selection of data for the CD-ROM is determined or guided by the market analysis, in particular by the selection of the user base. If the data are destined for a select group of persons who were involved in the collection and production of the data, the data that are placed on the CD-ROM may be all of the data that were collected and produced. If the data are destined for the larger user community, a carefully selected subset of useful data may be more appropriate for publication on CD-ROM. In either case, it is best to consult with persons who know the data that were collected, with persons who recognize the significance of this particular set of data, and with persons who understand the needs of the target user community. In some cases, these people may be the same set of persons.

Meetings should be held with members of these groups of people to decide what data are relevant for publication. Criteria for the selection of data should be laid out and data satisfying these criteria should be selected. The nature, size and format of the selected files plus any necessary ancillary files, such as calibration data and documentation files, should be determined. Any processing that has to be done to generate the products should be specified. For large datasets, such as large format imagery, the possibility of the compression of data prior to publishing, which can result in 50% space savings using lossless techniques, can be explored. The costs here should also be calculated, namely the location and testing of a suitable compression technique for the data in question and the ability of users to de-compress using software (for several potential platforms) on the disc.

These meetings should also be attended and tempered by persons knowledgeable about the budget available for data publication and in the data preparation and mastering costs involved in producing CD-ROMs. The time (and therefore cost) that is required to generate the products as selected by the knowledgeable persons should be estimated. If the amount of data or cost of

<div align="center">13</div>

producing the data exceed the budget, the knowledgeable persons must restrict their criteria further to reduce the cost. The meetings should produce a useful selection of data that can be processed and published at a cost that is within the budget.

Partly due to the general target user community, the selection of OTTER data for publication was a long (several months) procedure. From the 15 gigabyte suite of OTTER data, investigators had to select those images that would both represent the project and be useful for other scientists and educators. It was decided to publish data for all six sites at as many times as possible throughout the growing season and to publish all bands for data that were selected. For certain types of data that had special directional imaging characteristics, considerations such as time of flight and relative position of the sun during the flight became the driving criteria for image selection. All text data (meteorology, timber, etc.) that were available and deemed correct by investigators were published.

Section 7   CD-ROM DESIGN

7.1   Introduction

Once all of the data files have been identified, their location on the directory structure of the CD-ROM and the file naming convention must be decided. While CD-ROMs published by other groups can be used as a reference, using these structures and conventions for another CD may not be appropriate. Beyond the general rules which must be followed to conform to CD-ROM standards, other guidelines are offered in this section to aid in the actual design of the structure of the CD-ROM.

As with the determination of standards, it is best for one person familiar with the data to perform the CD-ROM design, rather than allowing a committee to come up with a structure.

Ideally, the data selection step should be completed prior to the CD-ROM design step. With all of the data to be published in hand, CD-ROMs with the best structure can be generated. However, particularly with scientists providing data, it is rare that all data are in hand prior to CD-ROM design. Many times it is necessary to proceed making the decisions faced in this section while data is being received. This may necessitate changing some of the decisions already made once the characteristics of the new data become known. This is a necessary evil, unless an unlimited amount of time is available to wait for data to be submitted by investigators. In the OTTER data publication, for example, the introduction of some video data (well after the file naming structure was defined) necessitated the definition of an exception to the file naming scheme.

7.2   Rules

Some rules, stipulated in the earlier design evaluation stage, will have to be observed in laying out the directory structure. For example, the file names must be DOS-compatible and the number of levels in the directory structure is restricted to eight for ISO 9660 implementations.

7.3   Directory Structure

The manner in which the files are laid out in the directory structure is dependent upon the organization of the data and the manner in which the target customer expects or can use the data. Typical organizations of scientific data on CD-ROM often reflect the location, date/time, and type of data collected, which is often the structure that is best understood by persons not familiar with the data. This may be difficult if several large format images are involved which together exceed the size of one disc. Looking at the structure of other similar CD-ROMs can provide

ideas about the proper directory structure for the data.

For the OTTER project, since the type of data is an effective means of subdividing the data, the top level of directories on the OTTER CD-ROM are titled with names of datasets. Since "site" is used heavily in OTTER scientific work, this was used as the second level of data categorization. "Site" could not have been used as the top level because there are some data files which cover more than one site or the entire state of Oregon. For those data files, such as AVHRR, "site" is not given as the second level of the directory. "Month of data collection" was used as the third level because of its importance in the OTTER project and in ecosystem research in general. Where there were multiple data files for one month, further (the fourth level) sub-directories were added to handle the day of the month in which the data were collected. Month of data collection is used as the second directory level for the AVHRR dataset. In the appendix is the directory structure of the first OTTER disc.

7.3.1   Multiple Disc Publishing Considerations

If more than one disc must be produced simultaneously, care must be taken to structure the directories such that similar data (data of the same type, location or date) will be on the same disc. This may be difficult if several large format images are involved which together exceed the size of one disc.

For the convenience of the user, directories for appropriate documentation and software should be on all discs. Documentation for data should be on all discs on which the data reside, even if the same file is repeated on more than one disc. The same applies to software which can be used on multiple discs. For example, if displayable images are to go on a particular disc, display software should be available on that disc, if display software is offered at all. In addition, information about important features on the disc should be provided, such as the precise location of sites within imagery on the disc.

In addition, a general file, entitled something like AAREADME.TXT, should be on all discs of a set. This file may be the same as the file on the other discs, if the file offers all information for all data on all of the discs. Or it may be a unique file which describes only the files on the disc. It is acceptable and easier to create one file with information about all discs and duplicate it on each disc.

7.4   File Naming

7.4.1   Introduction

A CD-ROM containing scientific data can consist of thousands of image, tabular, and text files. To help a user distinguish between files, which could vary by date/time or location collected or data type or other variable, the disc designer has two basic techniques: 1) use many directories and subdirectories named by variable to differentiate between files by location, or 2) select file names which uniquely identify each file. In practice, a combination of these techniques is used to differentiate between files. The technique of creating an appropriate directory structure was discussed above. To differentiate completely by directory structure, however, would result in more directories than is reasonable and in excessive directory traversals by the user.

A file naming convention must be set up to distinguish between the files on the disc. One might say that it is only necessary to distinguish between the files in a single (sub)directory, since the directory structure identifies the other characteristics of the files. This is true, however, doing so may result in file names which are the same as other directories. For example, a file name of 120000.IMG (for an image taken by one sensor at 12 noon: 12 hours, 00 minutes, and 00 seconds) may be identical to a file name of a different sensor (in a different subdirectory) which

took data at exactly the same time. One might also say that it doesn't matter that the file names are the same as long as the files are on different directories. If, however, the user transfers both files to the same directory on his/her own disk, there will be confusion unless the file names are different. For this reason, it is advised that a policy be adopted requiring that the file name for each file on the CD-ROM be unique.

7.4.2 Unique File Names

Unique file names across the entire disc greatly complicates the naming of files and usually has the disadvantage of creating cryptic file names with no real meaning for the user. It is desirable to make file names that reflect the contents of the file. With the ISO9660 standard of a file name consisting of no more than eight characters followed by a three character extension, there is little room for meaningful information. Particularly with files that are differentiated by several variables (e.g., data type, date/time, location, channel number), little meaningful information can be stored in eight characters.

The technique most often used is to code information into one, two, or three characters and then assign each code a position in the file name. For example, data types can be abbreviated to somewhat meaningful two-character strings (e.g., AS= ASAS and AV= AVIRIS) and used as the first two characters of the file name. If data were collected over only a few dates, perhaps a one letter code can be used to represent each date and be the third character of the file name. Site names can be reduced to a number or to a character.

When designing the file naming scheme, care should be taken to consider all of the types of files that will be on the disc. While all files (such as documentation files) don't have to strictly follow the file naming convention (e.g., documentation does not have to begin with a two character "data type" designator, such as DO), the scheme should be understandable and require a minimum number of exceptions. In the OTTER file naming convention, three characters were reserved for the image band number, of the form: Bxx, where B is constant and xx is a number from 01 to 99. When naming the meteorological data files, which of course is not imagery and does not involve bands, the three character slots (in combination with other slots) are used to describe the frequency of data collection, HOURLY or DAILY. Exceptions such as this should be avoided where possible and, when done, should attempt to make the file name more representative of the contents of the file.

This scheme works if there are not too many pieces of information that must be recorded in the file name. This is often the case when the time of a data acquisition must be recorded. Two characters each for hour, minute, and second leaves little room for other information. If there are too many pieces, a coding scheme may be adopted for all or part of the file name. This code for the file must then be recorded in documentation on the disc for reference by the user. The file name does not bear any (or much) resemblance to the contents of the file, but the unique file name can be preserved.

7.4.3 Conclusion

The standard which was selected for the presentation of the data on the CD-ROM will undoubtedly have an effect on the file naming convention. For example, the three character extension is dictated by the PDS standard, so those characters cannot be used

The file naming scheme should be developed by one person who should take several weeks to ponder the advantages and disadvantages of several file naming options. The candidate schemes should be reviewed for ease of use and understandability by others familiar with the data and familiar with the use of CD-ROM.

The file naming scheme adopted for the OTTER data is shown in the appendices.

Section 8   FILE PREPARATION

8.1   Introduction

File preparation involves creating and processing the data and documentation (selected in the "Data Selection" section) and software files to the standards and design adopted in the "CD-ROM Design Considerations" section. This is potentially one of the most time consuming tasks, reflecting the number of files involved and the amount of processing that can be undertaken for each file. Data preparation also typically involves creating files which serve, or increase the understanding of, the main data files.

It is recommended to have all data, documentation and software prepared and validated before designing the disc, so that file sizes will be known and the nature of the data to be put on the disc can be identified. It allows more time to become familiar with the data and consider different possible directory structures and file naming schemes before deciding on a final configuration. In practice, however, since the data and document preparation can involve considerable waiting for some data providers to provide their data, time is available to plan the structure of the directories and file names and to write general and documentation files. As long as nothing is considered final, such "jumping ahead" causes no harm. When the data does come in, changes may have to be made to the structure and documentation files (not to mention the CD-ROM artwork and booklet).

Performing data preparation first also helps to identify problems with the data that may result in the removal or substitution of data and that will have an effect on the disc design. Although this should be uncovered in the data selection step, the need to compress large format data sets may also become obvious if it is discovered that more discs are required to publish the data than the budget allows.

A Data Preparation Plan should be written to make sure that all selected files are being processed and to be able to schedule the preparation of the data.

8.2   Large Versus Small Amount of Files

If the number of files is few or the changes are minor, each file can be handled individually. If, for example, files are to be placed on the disc in their native format, only creation of ancillary and documentation files may be necessary.

If the number of files and the number of changes that have to be done are great for any type of file, then a means of automatically producing the required files becomes more attractive. Saving time and cost in the long run, this may involve writing software or command files to generate these files. Files in the proper format and named according to the file naming convention can be the output from the software.

The accompanying ancillary files can be documentation files, ancillary data files, data format files, and files which may facilitate the use of the data files, such as PDS label files. If the number of files and calculations are complex, it may be best to incorporate the creation of these files as part of the software written in the above paragraph.

Whatever the technique for creating the new files, the procedure should be tested on one data set to make sure the process is working properly. If multiple stages of processing are involved,

testing of the results of each stage should be performed. Only when the final product is correct should all files be processed.

8.3   Binary Data

Different types of files (e.g., binary versus numerical tables) typically require different types of processing. Image data requires manipulation of the imagery in its given form to a form selected in the "CD-ROM Design Considerations" section. If it is decided to publish the imagery in its native form, then little needs to be done besides carefully copying the data from the original media to the proper reserved directory location on the disk and producing any necessary ancillary files. If a data presentation scheme has been selected, then descriptive information must be put in the format of the presentation scheme. For example in the OTTER case, PDS label files must be generated for each image file, using the fields selected in the Design section.

Creation of the files to be published can utilize commercial products or newly-developed software involving several manipulation steps. Well before actual processing, the procedure should be tested on one data set and, since considerable processing is likely to be performed, a schedule for all data processing should be written.

For the first OTTER CD-ROM, software was written to perform two basic tasks: 1) generate PDS labels for each binary image file, and 2) create image files that can be displayed to verify data quality.

8.4   Text Information and Data

Text information and data includes documentation, tabular data, formats, ancillary data- basically, anything that is expressed in ASCII format.

8.4.1   General Disc Documentation

General documentation that helps the user to best use the disc must be written for publication on the CD-ROM. The information that should be relayed to the CD-ROM user includes:

Overview of the project which created the data and/or the reason for creating the disc
Overview of the data (data types, quantity, and other characteristics) that are on the disc
Organization of the CD-ROM
Overview of the other documentation files (names and location) on the disc OR format of this file (if all general documentation is in this file)
Hardware/Software instructions for the reading of the disc
A name and phone number of someone to contact for assistance with the disc
A name and phone number of someone to contact for assistance with the data
Acknowledgment of the assistance and support of others

Directory structure of the disc
File naming convention
Index of data files (particularly if several files and/or the file name does not reflect contents)
Detailed information on the project or the techniques for generating the data set
Specialized information pertinent to the data
Glossary of terms, keywords, and acronyms

The reason for dividing the list into two sections is to highlight the option to place all of the information in one file or to place some of the information (the items above the blank line) in an introductory file and the remaining items in individual files that are referenced in the

18

introductory file. For general readability and ease-of-use, it is recommended to do the latter with a limited number of relatively small files. One specific advantage is that the user is able to easily print out individually any aspects which they may like to have available for easy reference, such as the directory structure.

The file name for the all-inclusive file or the introductory file typically includes the string "README", which is a standard for PC software documentation. The file can almost assured to be the first file listed in an alphabetical listing of files with the name, "AAREADME.TXT". Other variations including README.1ST, README.TXT and README.DOC, can be used. The extension will vary depending upon the file naming scheme selected.

There are no standards for the structure of the README and the other general documentation files. In the validation stage to follow, only clarity and completeness of the files will be tested.

The OTTER AAREADME.TXT file is included in the appendices. The directory structure, referenced in the AAREADME.TXT file and labeled DIRTREE.TXT on the CD-ROM, is also included. A specialized document, SITELOC.TXT, was also generated and published to help the user to locate the OTTER sites both on maps and on the imagery on the disc.

8.4.2 Data Documentation

Separate documentation files are typically written for each of type of data and applies to all data sets within the data type regardless of location on disc. For each type of data, these files can describe how the data were generated, what the data were used for (if generated by a project), what data were collected (in general terms), and who to contact for information on the data, and lists references to other related documents.

Format documentation is typically created for tabular data using a word processor because, while the total number of data files may be large, there are typically only a few different formats in which the data are placed.

PDS label files can be generated by using a word processor (advisable if publishing only a few such files), or by executing a label generation program written by PDS staff members, or by writing software or command languages to generate the files.

Example OTTER data type documentation (for ASAS data), format (for meteorological data) and PDS label files (NS001 image data) are included in the appendices.

8.4.3 Tabular Data

Tabular data files could be voluminous and their creation or re-formatting could require considerable processing. Automated techniques utilizing operating system command languages or are generally recommended over third generation programming languages for reasons of simplicity and maintainability. If a tabular file of rows and columns can be manipulated to the chosen format using very high level field shifting commands, the work level is reduced considerably. Large numbers of identical changes to a few files can be accomplished using sophisticated word processors. Considerable time could be saved even for a few small files consisting of ten or fewer rows and columns. Manual re-formatting of tabular files should generally be avoided.

Several staff-written Unix C-Shell scripts were used to process the OTTER tabular data to a proper format. At times, data extracted from a database was manipulated by such scripts.

8.5   Software

19

Based on the platforms and software packages selected in the "CD-ROM Design Considerations" section, files of public domain software should be prepared for publishing.

It is advised that no recommendations to use specific commercial programs (Photoshop, etc.) should be made. It is appropriate to just note that the files are in a generic format which can be read by many commercial applications.

8.6   Placement of Files In Directories

As the data and software files are prepared, they should be placed in a directory structure that has been reserved for CD-ROM files on a hard disk. This directory structure was defined in the CD-ROM Construction step and emulates what will eventually be written to the CD-ROM. The placement of the files can be done from software which creates and writes files to the specified directory. Or it can be done by manually copying files from personal workspaces to the proper directories. It must, therefore, be assured that there is access to sufficient contiguous disk space on the computer to hold the amount of data that will be written to the CD-ROM, which could be 650 (or more for multiple discs) megabytes of data.

Section 9   FILE VALIDATION

9.1   Introduction

After preparing the data and placing the files in the CD-ROM directory structure, the data and documentation files must be validated for completeness, accuracy, and usability. A manager in conjunction with the staff should develop a test plan which checks each type of data against these criteria. Due to basic differences in the structure and purpose, each type of data requires a separate test plan. The test plan should be implemented by someone who did not do the original processing but is familiar with the data and the target market and has a "critical eye". After a critical inspection by a staff member, it is also beneficial to allow actual target recipients of the disc to look at the second version of the files.

9.2   Validation Criteria

A test plan should be written to include a complete checklist of criteria to validate. Because the data for each data type is unique and the uses of the data varies by project, it is difficult to lay out criteria that should necessarily be validated. The following are incomplete guidelines that apply to most data types.

All files necessary to understand the data must be present: format, documentation, calibration, header, PDS label, etc. The documentation must be complete and clear enough to allow users to apply the data in common scientific endeavors. The data values in the files that have been re-formatted must resemble the original data values. Format files must be checked to make sure they reflect the actual data files. Any data transformations must be validated by other means than the original means (software or whatever). The software on the disc must process the data successfully on the target platforms with no (or at least documented) errors.

If data are changed to correct detected errors, then these newly-generated files must be checked against the same criteria.

There are many other criteria which could be evaluated, but we must also warn against excessive testing. All errors uncovered in "reasonable" checks of the data should be removed, but trying to please the perfectionist user will add greatly to the costs of CD-ROM production. Follow the

simple tenet, adding "within reason" at each adjective: as long as the data are complete, accurate, and usable, the data pass the test. There will be a version number on the disc, meaning that changes or errors uncovered after initial mastering can be corrected.

Because the task of image subsetting was a major portion of the processing performed on the OTTER aircraft imagery, a test plan was developed to check the performance of the software. This test plan is offered in the appendices as an example.

Section 10   PROOF DISC PREMASTERING/MASTERING

10.1   Procedural Overview

Once the data, documentation, and software files have been prepared and validated, the procedure of creating the CD-ROMs can begin. This process involves premastering, creating and testing proof discs, and eventually generating the final CD-ROMs. It is a somewhat complicated and potentially time-consuming process which offers some hardware/software configuration options along the way that affect the cost of the work. Here is an overview of the procedure.

The first step of the procedure, which is so highly recommended as to be a requirement, is to create some proof, or one-off, discs for testing purposes. A proof disc is a temporary copy of the data is cheaper than the entire mastering process of creating the glass master and generating replicates. There is no artwork on the proof disc, but it contains, in CD-ROM format, all of the data that was written to the reserved directory.

To create proof discs, it is necessary to premaster the data resident on the reserved disk directory. Premastering is the process of converting the prepared data into a form that can be written onto discs. Premastering is performed by software which can write the converted data to hard disk, magnetic tape or directly to the proof disc, depending upon the hardware/software configuration available. Through differing mechanisms, which will be outlined in the following subsection, the proof disc is generated.

After the proof disc is reviewed and required changes are made, either a second proof disc can be generated from the revised data or the final mastering can be performed. If the second proof disc is generated, a review would again take place resulting in another version. More proofs and versions could be generated until confidence is reached that the discs are "complete, accurate, and usable". When the decision is made to do the final mastering, the final version of the data (along with booklets, inlay trays, and disk art) are sent to the mastering facility which creates the master disc and all of the replicates packaged in plastic "jewel boxes" covered in shrink wrap cellophane plastic.

There are three basic ways in which this procedure can be accomplished, each with their own costs and advantages and disadvantages.

10.1.1 Outside Premastering Service

The procedure to use an outside premastering service to produce the CD-ROMs requires the purchase of no in-house software or hardware for CD-ROM production purposes and follows these steps:

1) Arrangements are made with the purchasing department to purchase the services of a premastering service. It may take months or hours for the paperwork to work its way through the system, depending upon the institution involved.

21

2) The prepared data may be written from the disk directory structure (which emulates the structure as designed for the CD-ROM) to a magnetic tape using a utility which should be compatible with utilities available at the premastering service. It is preferable to be able to write all of the data on the directory structure to one tape. Determine what media is acceptable by the premastering service. After premastering the data to tape, it is recommended to back up the data directories using an operating system utility, such as the Unix operating system's "tar" command.

3) The premastering service will take the data tape and create a one-off disc which will be sent back for disc review (see below).

4) After changes are made to the original disk files in response to the review, a new magnetic tape is written and sent to the premastering service. Disc artwork and a booklet and inlay tray, if created, is also sent to the premastering service.

5) The premastering service premasters the data to a one-off disc and sends the disc and the artwork to the CD-ROM mastering facility, which creates the master disc and replicates.

All of the costs are paid to the premastering service, which pays the mastering facility for its services. This option is usually advisable if only a couple of discs are involved and no more are anticipated in the foreseeable future.

10.1.2 Inside Premastering/Outside Mastering Service

The procedure to purchase software to perform the premastering step and use an outside mastering service to produce the CD-ROMs follows these steps:

1) Make arrangements with the purchasing department to purchase the premastering software and the services of a mastering service (for one-offs and final replication). It may take months or hours for the paperwork to work its way through the system, depending upon the institution involved.

2) The prepared data may be written from the disk directory structure (which emulates the structure as designed for the CD-ROM) to a magnetic tape. It is preferable to be able to write all of the data on the directory structure to one tape. Determine what media is acceptable by the premastering service. After premastering the data to tape, it is recommended to back up the data directories using an operating system utility, such as the Unix operating system's "tar" command.

3) The mastering service will take the data tape and create a one-off disc which will be sent back for disc review (see below).

4) After changes are made to the original disk files in response to the review, a new magnetic tape is written and sent to the mastering service. Disc artwork and a booklet and inlay tray, if created, is also sent to the mastering service.

5) The mastering service creates the master disc and as many replicates as were ordered and sends them back.

This option is usually advisable if more than a couple of discs will be produced.

This was the technique used in the publication of OTTER data. The Makedisc CD-ROM premastering software was purchased and installed on the system (Sun) on which the data to be published was stored. Using the Makedisc software, the prepared data was written from the directory structure and stored on Exabyte type. The tape was sent to the mastering facility and

two one-offs were returned. After review by investigators and other reviewers, the Makedisc software was executed again with the corrected data. The disc artwork, the printed booklet, and the inlay tray was sent with this final version to the mastering facility, which returned the shrink-wrapped replicates. The purchase request for the premastering software and for the mastering services is given in the appendices.

10.1.3 Inside Premastering, One-Offs/Outside Mastering Service

The most sophisticated procedure available to date is to purchase hardware and software to perform the premastering step and to create draft CD-ROMs (one-offs) in house and then use an outside mastering service to produce the final CD-ROM masters and replicates. Because of the decreasing prices of the machines which produce the one-off discs (which cost approximately $35 each), the OTTER project would purchase one of these devices instead of just the software, if it were purchasing as of this writing. To aid in the selection of machines for one-off production, contact the JPL Data Distribution Laboratory. The procedure generally follows these steps:

1) Make arrangements with the purchasing department to purchase the premastering hardware and software and the services of a mastering service (for mastering and replication). It may take months or hours for the paperwork to work its way through the system, depending upon the institution involved.

2) The prepared data may be written from the disk directory structure (which emulates the structure as designed for the CD-ROM) directly to a one-off disc. As many discs that are necessary for review (see below) can be generated in house.

3) After changes are made to the original disk files in response to the review, a new one-off disc is written and sent to the premastering service. Disc artwork and a booklet and inlay tray, if created, is also sent to the premastering service.

4) The mastering service creates the master disc and as many replicates as were ordered and sends them back.

This option is usually advisable if several discs are being produced and more are anticipated in the foreseeable future.

Section 11   PROOF DISC REVIEW

11.1   OTTER Proof Disc Review (Coordinator Description)

The first task upon return of the proof (one-off) disc is for the staff to validate various characteristics of the disc. Using the test plans, selected criteria can be evaluated. And don't overlook the simple characteristics. For example, try to read each type of file into common software on each target platform and perform basic commands (such as "type" on DOS) on random ASCII files. These kinds of reviews require access to CD-ROM drives on the target platforms. Unless you plan to use the drives for other purposes, it is unnecessary to purchase the drives just for testing. It is adequate to borrow a colleague's drive, if available.

The major review of the proof disc is performed by a team of reviewers. This group should consist of project members familiar with the data, person(s) familiar with the publication of similar data on CD-ROM, and outside scientists of the same discipline as the project scientists. With planning and some luck, both the concerns of the persons who provided the data and the needs of the target market should be represented. Since there can be many reviewers, it may be

wise to create at least three proof discs so that the review process can be expedited. It is possible for one disc to be sent between reviewers, as well. If the reviewers are located together, the team could review the disc together. If possible, it is desired for all reviewers to meet to discuss their individual impressions of the disc, either in person or via conference call. Perhaps a set of prioritized recommendations can be a result of the meeting.

It is advised to take all comments in balance with others and "with a grain of salt". Individual reviewers may be seeing the disc from a somewhat distorted point of view and not represent the general market. If several reviewers express the same sentiment, the comment should be taken much more seriously. Remember that there is a wide spectrum of potential users for the disc (the "silent majority", to borrow a phrase from the 70's). Distinguish between correctness of data and style preferences. Make a decision whether to modify the files based on the target market and whether their interests will be advanced by the change. If a notable advantage will be gained and it is within the budget to implement, make the change. If the advantage is questionable or minimal, more harm than good may be done by the change. Errors of commission or omission, of course, must be corrected. It is difficult to estimate the amount of time (and therefore money) that will be necessary to correct the data, but a thorough market analysis at the beginning of the project will minimize these expenses.

Since you earlier backed up the raw files on the one-off, you can change the files on the directory structure in preparation for the production of another one-off or the final mastering.

The first OTTER disc was reviewed by all OTTER investigators (and/or their graduate students) who had contributed data for publication. No outside members representing the target market were tapped to review the disc. One comment that the structure of the files should be left in their native formats was rejected because of the desire to make the discs as generally useful as possible. The easy-to-use formats of the files on the CD-ROM overshadowed the argument that some software routines, written for the native formats, will no longer work.

For the remaining OTTER CD-ROMs, two members of the scientific community who are familiar with the target market will also be reviewing the discs.

11.2   OTTER Proof Disc Review (Liaison Description)

This section contains procedures and observations of the review of the first OTTER CD-ROM by J.W. Skiles (of Johnson Controls World Services), who was the scientific liaison to the reviewers:

The OTTER Project involved over 20 investigators and collaborators. Given such a pool of interest and expertise, I expected that six persons could be found who would volunteer to review the first OTTER CD-ROM. After the 8mm tape containing the pre-mastered files for the CD was sent to DMI, I contacted by telephone eight OTTER investigators and asked them to review the contents of a proof of the CD. Of those eight, only six agreed to act as reviewers. In order to get their cooperation and to save time, I proposed that I send the disc out to the first person who would then look it over and pass it on to the next reviewer who would then do the same and so on until each person had seen the disc. I had to promise that the postage for sending the disc on to the next reviewer would be supplied and that the envelopes would already be addressed. This way the reviewer would receive the disc, review it, and send it out with just the expenditure of time as the cost of each review. I also extracted a promise from each person that s/he would take no more than three working days to review the disc.

To facilitate my tracking of the disc during the review process, I enclosed with the disc and the postage-paid envelops, a pre-addressed postcard (one for each reviewer), with the NASA frank, and requested in a cover letter to each reviewer that the card be sent back to me at the same time

24

the disc was sent to the next person on the distribution list. My cover letter also included my e-mail address with the explanation that it could be used to send me the status of the disc instead of the postcard and that even the reviews could be sent to me this way instead of using a formal letter. This review process should have taken no more than three weeks, if one includes the time the package was in the mail.

Concurrently, a second review copy of the disc to a program manager at NASA Headquarters for review. A third proof disc was kept at ARC for the resident OTTER investigators to review. It was displayed in my office and a checklist of persons who checked it out, reviewed it, and returned it to my office was written on my blackboard.

11.2.1 Outside Peer Review

The outside peer review took more than two months. People who had promised to review the disc contents and get the disc into the mail in three working days ended up taking weeks. The reasons varied: One investigator gave the job to a graduate student who did not realize the time constraint; one reviewer was away to a conference for a week when the disc got there, and he did not come across the package in his accumulated mail for more than a week after he got back; a third reviewer had managed to break both of the CD readers at his disposal just prior to the time the package got to him. He waited for more than a week before telling me of the problem and he then sent the disc on without reviewing it.

Another delay came about because two of the reviewers are Canadians and the package had to pass through customs each time it crossed the international border between the United States and Canada.

11.2.2 Headquarters Review

The disc sent to a headquarters manager disappeared and was never passed on to other personnel at HQ as requested. No comments on the proof disc ever came back from NASA HQ.

11.2.3 ARC Peer Review

No one at ARC took or had the time to look at the proof disc. One investigator asked to view several Daedalus scenes from the disc on a PC. Another asked that one of the ASCII files on the disc be printed out for review. No other reviews were done at ARC.

The comments we did get from those persons who did review the first proof disc necessitated a second proof disc. Two were produced. I sent one copy to two OTTER investigators, one who had seen the first proof disc, and a second who had not reviewed the first disc. The disc was sent out the same way

11.2.4 Observations

The review process of the proof discs was hampered by the lack of funds to support what was essentially a data management job. That is, most investigators were not funded to review the accumulated data. Further, most were not conversant with all of the types of data and imagery on the disc and they could not therefore render an evaluation about the total contents of the disc.

The review of the discs is very much different from reviewing a paper for journal publication. A journal paper is usually confined to the reviewer's area of expertise and consists of written pages with figures that can be marked and returned to the editor. The OTTER disc contains more than 566 megabytes of data and images, neither of which can be marked or corrected without first transferring them to a computer and then printing them out. The data on the disc are varied as

they include timber and canopy chemistry data and imagery from seven instruments. Thus, the data cover many areas of expertise.

The real reason for the difficulty of getting the disc reviewed is that there was nothing in it for the reviewers. Their names did not appear as reviewers and they were not paid to offer their opinions. Hence, they were not motivated to aid in the review process.
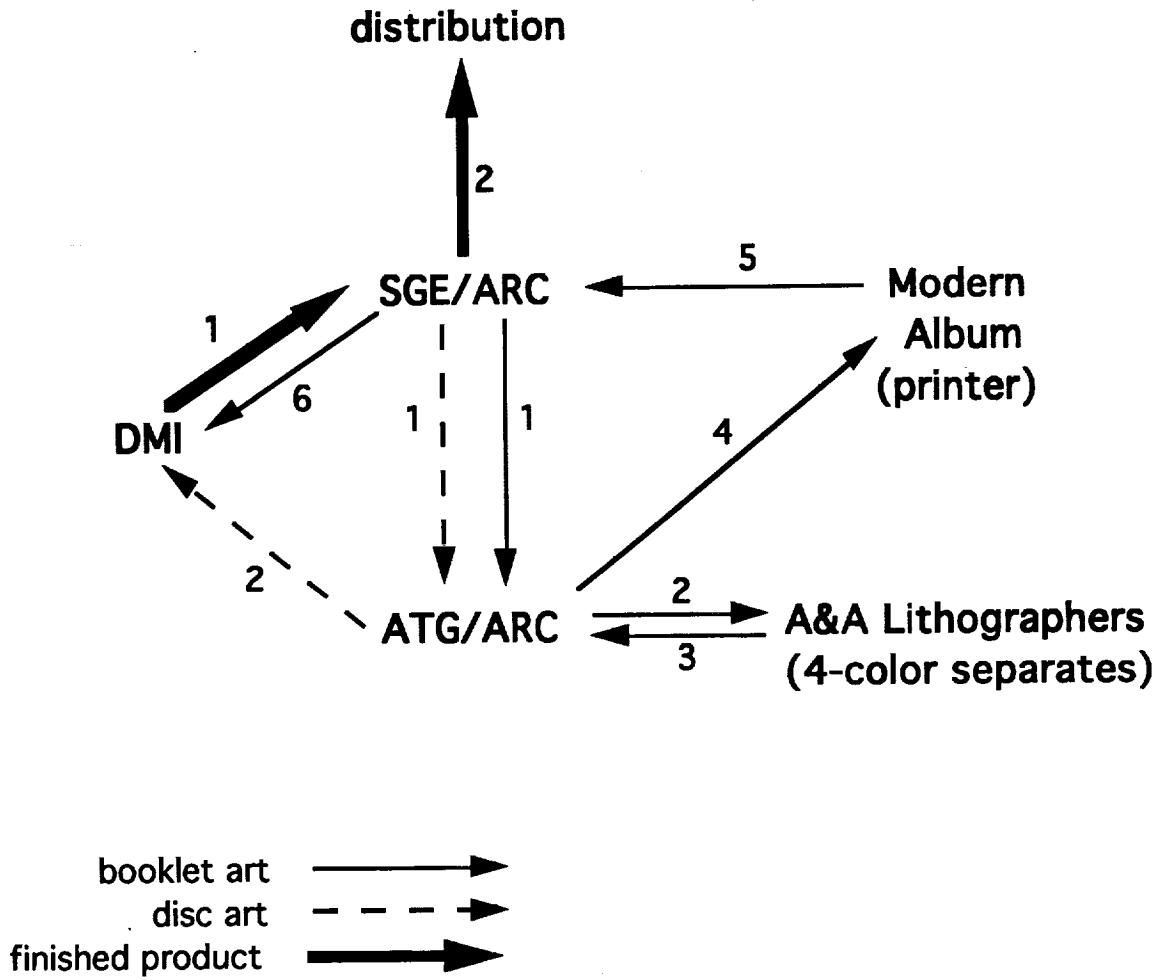
Section 12   DISC AND INSERT ARTWORK DESIGN AND PRODUCTION

12.1   Introduction

This section was written (with modifications by the author) by J.W. Skiles, a Johnson Controls World Services (JCWS) employee who was responsible for the production of the artwork, booklet and inlay tray for the first OTTER CD-ROM. He is thanked for his contribution to this Sterling Technical Note. This section describes the experience of producing these important products which are the first things that are seen by the user of the OTTER disc. The reader should be able to glean some lessons that were learned and apply them to his/her own production of artwork and inserts.

The people at Ames Research Center who were involved in the design of the artwork used in the production of the OTTER CD-ROMs are Merin McDonell (Johnson Controls World Services, Code SGE), David Faust (Quad S, Code ATG), and J.W. Skiles (JCWS, SGE).

The entire process of disc and insert artwork design and production, which can be referenced during the description below, is represented pictorially with the following chart:

distribution

SGE/ARC

Modern Album (printer)

DMI

ATG/ARC

A&A Lithographers (4-color separates)

1

6

2

1

1

2

5

4

2

3

booklet art ———————▶
disc art – – – – ▶
finished product ━━━▶

This following table shows the major steps involved in the production of the OTTER CD-ROMs:

Production Step	Where Performed	Personnel	Dates	Software/Equipment
Design Specification	SGE/ARC	Angelici, McDonell, Skiles	Aug 92	paper, colored pens & pencils
Initial Booklet Cover Design	SGE/ARC	McDonell, Skiles	Aug-Sep 92	Aldus Freehand, Adobe Illustrator, Macintosh
Initial Disc Art Design	SGE/ARC	McDonell, Skiles	Aug-Sep 92	Aldus Freehand, Adobe Illustrator, Macintosh, scanner
Initial Inlay Tray Design	SGE/ARC	McDonell, Skiles	Aug-Sep 92	Aldus Freehand, Adobe Illustrator, Macintosh
Initial Booklet Design	SGE/ARC	Skiles, Angelici	Aug-Sep 92	Microsoft Word, Macintosh
Reworked Cover Design	ATG&SGE/ARC	Faust, Skiles, Gifford	Sep-Nov 92	Adobe Illustrator
Reworked Inlay Tray Design	ATG&SGE/ARC	Faust, Skiles, Gifford	Sep-Nov 92	Adobe Illustrator
Reworked Booklet	ATG&SGE/ARC	Faust, Skiles, Angelici	Sep 92	Adobe Illustrator, Adobe Pagemaker, Macintosh
Two-Color Separates (disc)	ATG/ARC	ATG	Oct 92	
Four-Color Separates (booklet & inlay tray)	A&A Lithographers	A&A Lithographers	Nov 92	
Premastering	SGE/ARC	Angelici, Popovici	Sep 92, Dec 92, Feb 93	Makedisc Premastering S/W, Sun computer
Proofs	JPL Data Distrib. Lab., Disc Manufacturing, Inc.	Angelici, Skiles, Popovici, DMI, Sorensen (JPL)	Dec 92	Macintosh, PC
Printing	Modern Album	Jack Bratel	Nov 92	
Production	DMI	DMI	Feb 93	

12.2   Disc Artwork Design

Artwork for the front of the CD was constrained by funding and the number of logos and emblems required on the disc. The funding limited the number of colors that could be used on the disc face to two. A higher level of funding would have allowed for more colors and would have made the disc art more striking and interesting. It was necessary to put on the disc face the "COMPACT disc" logo, the "ISO 9660", and a logo for the Pilot Land Data System. Each of the logos had to be large enough to distinguish the lettering and lines; this was especially a problem with the PLDS logo.

There was also a need to uniquely number the disc. The string "USA_NASA_PLDS_OT_00X" was used, where X denotes a volume number with a "1" for the first disc.

All of the above had to be balanced in the limited space of the disc face. A design for a solid, 3-D impression of the state of Oregon on the left side of the disc was used to offset the lumped logos that we put on the right center of the disc. McDonell varied the stippling of the stylized trees for part of the design which gave the impression that several shades of green were used.

The disc art as done by McDonell needed no re-working. The color negatives for the disc art were done by the Ames Graphics Branch. These consisted of two negatives, co-registered so that the objects on each negative matched. One negative showed the black portion of the disc design while the other showed the green part of the design (see Appendices). The Graphics Branch did not charge for this first experience of CD-ROM production.

The disc art negatives were sent to the mastering facility, Disc Manufacturing, Inc. (DMI) in Anaheim, California on October 27, 1992 with the specification that the green color used on the disc face be Pantone scale #370. The Graphics Branch paid for postage for this mailing.

Black and white copies of the OTTER CD-ROM disc art are included in the appendices.

12.3   Booklet and Inlay Tray Artwork Design

In August, 1992, McDonell and Skiles began designing the cover for the booklet and the inlay tray. McDonell came up with several designs that she and Skiles re-worked with an eye to making the product visually appealing and distinctive. As part of the design, the NASA "worm" logo was used on the cover of the booklet and on the inlay tray (it was used on the face of the disc as well). This was about the time that the NASA Administrator decided to make the older "meatball" the official logo for NASA.

The meatball logo is a distinctive sigil that features the lettering "National Aeronautics and Space Administration" with a stylized rocket ship orbiting a moon all on a backdrop of blue sprinkled with white stars. It is a very busy logo and probably has some historical significance, but in order to make all the parts of it legible on the CD artwork, it would have had to be the size of a quarter and it would have dominated the design on the disc and the booklet.

At this time too, we found a statement in the document by Blanche Meeson on how to produce a CD-ROM that said all CD artwork had to be passed through an official in Code LPS at headquarters; we were to allow up to six weeks for the approval process.

We thought at the time that we did not have the six weeks to wait for approval. At this point, I contacted the Imaging Technology Branch and was told the design had to be approved by NASA Headquarters before they would look at it. I then contacted David Faust (copy artist) and Loren Gifford (Chief) of ARC's Graphics Branch (Code ATG). I showed them the designs that McDonell had developed for the disc front and for the inlay tray and booklet, all of which used the NASA worm logo. They assured me that they had discretionary authority to use the worm and that they felt justified in doing so since the design was done before the pronouncement from Goldin.

In order to accommodate all the logos, content lists, and designs on the inlay tray and booklet, Faust had to re-work some of the elements in the design by McDonell. Further, he had to produce a tray and booklet with the designs on top of a color photograph taken from a 35mm slide provided by an OTTER investigator. This process required several weeks for two reasons. The first was that he could not devote full time to this one project and the second was that we were not finished finalizing the contents of the disc for some time.

Faust also re-worked the text for the inside of the booklet, setting it up in Copperplate font and

positioning the pages so that when cut and folded at the printer the pages would be in order. (The pages of a booklet that fits in a CD-ROM jewel case are laid out four at a time on a single piece of film. The booklet for the first OTTER CD-ROM had 16 pages including the covers.)

Once the design of the booklet and inlay tray was finalized, the printing had to be done, but first color composed film or color separates had to be made. Upon the recommendation of Loren Gifford, I took the job to:

A & A Lithographers
3501 Haven Ave.
Menlo Park
415/365-1919

A four color booklet cover and inlay tray requires that four different separates be shot. The separates are done using the original artwork with color filters placed in front of the artwork and four pictures taken in light of colors specific to the primary colors used in color photography. The result is four sheets of black and white film negatives with the black on each negative corresponding to the light of three of the primary colors. When all four colors are merged, the result is a full color picture. The equipment and cameras necessary to make the separates is large and expensive; ARC does not have such a facility, so the job had to be done off-base. The final product was four sheets of film for the booklet cover and inlay tray as well as several single sheets of film, black on white, with the booklet text.

The lithographers did not produce the side panels of the inlay tray correctly the first time; they did not "bleed" the picture onto the panels, making the volume number standout as black on white on one side and white on white on the other side of the tray. I insisted that they shoot the inlay tray over after making the appropriate changes.

Gifford was (and still is) interested in promoting the use of CDs as a medium for storing large volumes of data. The cost of producing the color separates ($610 plus California sales tax) was borne by the graphics branch to facilitate our working with them in the production of the first OTTER CD-ROM.

The production of the four color separates by A & A Lithographers was completed during the second week of November, 1992. The graphics branch at ARC sent the separates and the booklet text to Modern Album (see below) on November 20, 1992.

12.4   Booklet Printing

The bids for printing a booklet were solicited from several printers over the telephone. The names of the printing companies were supplied on a list from DMI. The printers and their bids are as follows:

•Stoughton Printing
130 N. Sunset Ave.
City of Industry, CA 91744
213/686-2753
contact: Jerry Brown
specs - 1,000 minimum; $638; requires composed film
(revision; price for above changed to $999 for 12 page booklet plus inlay tray)

•Golden Rule Printing
1864 Sparkman Dr.
Huntsville, AL 35816
800/239-4653
contact: "Brad"
specs - 1,000 minimum; $750; requires 4-color composed film

•Modern Album
3116 Van Owen St.
Burbank, CA 91505
818/841-8683
contact: Jack Bratel
specs - 1,000 minimum; $770; requires composed film and color key; will send samples

A fourth printer in Alabama was contacted twice over the telephone, but never did give a price quote. Because Modern Album was in the same time zone as ARC and we could therefore talk with them on the telephone at the beginning of our day and until our closing time, I decided to use Modern Album rather than Golden Rule Printing. The $27 difference between the two was probably offset with the reduced price of shipping the product within the state as opposed to shipping from Alabama to Ames. Modern Album did send samples of their work which no one else was willing to do. This helped in visualizing how the OTTER booklet would look in the final form.

The cost for the printing of the booklet and inlay tray was $833.53 ($770.00 plus California sales tax of 8.25%). Modern Album returned the printed job via Crescent Truck Lines. The shipping charge was $39.13.

The booklets and inlay trays were sent to DMI in December, 1992. They had the artwork for the disc since October of 1992.

12.5   Production Errors

In producing the disc artwork at ARC, Faust has used two concentric circles as guides, one delimiting the outside of the disc and one for the hole in the center of the disc. Unfortunately, he made these guides black, so DMI thought they were part of the design. They were not. However, I got a panic call from DMI when they went into production on the OTTER job saying the circles were not exactly concentric, what should they do. I replied that they were to ignore the circles and proceed. This delayed the production by at least a full working day.

After all the work that was put into the production of the disc and accompanying material, two errors appeared on the OTTER CD-ROM. The first was that the lettering on the disc front is not symmetrical. This is not really noticeable, but it does detract from the overall esthetics of the disc. (I have been assured that this is a feature built into every font. That is, rounded capital letters are a very small amount bigger than straight-line letters. This is done because the eye would see the rounded letters as being smaller than the other letters otherwise.) There is no way to correct this in subsequent CD production and it is only noticeable to a few.

The second error came about because the silk-screening process used by DMI to produce the disc front does not use the Pantone scale of colors. As a result, they tried to match the green color I specified, but they were not entirely successful. The green they used on the disc front is more yellow than the Pantone #370 used on the front of the booklet. There is no simple fix for this problem as no bridge between the two color scales exists.

Section 13   FINAL DISC MASTERING

After the data on the reserved directories have been modified to reflect selected reviewer comments, the final premastered tape or one-off (with disk art, booklets and inlay trays, see section 10) is sent to the mastering facility. In the turnaround time requested in the purchase order, the mastering facility returns the ordered number of shrink-wrapped jewel boxes, ready for distribution (see section 11).

Section 14   FINAL DISC DISTRIBUTION

During the months before the final mastering of the disc, a list of prospective disc recipients should be constructed. Their addresses and phone numbers should be verified, and the information should be recorded in a database or spreadsheet (or even word processing) application. As the discs are mailed to recipients, record the date sent into the application. This information is important if subsequent discs or notifications must be sent.

The discs can be mailed using special CD-ROM mailers available from many office supply companies, or, in a pinch, jiffy packs or boxes with padding can be used. Be aware that mailing to foreign countries may require special approval in advance.

To advertise the CD-ROM's availability, several venues are possible and vary with the data being published. Papers, articles, or even two inch advertisements in journals read by those who could use the data in their research is a good place to start. Government publications, such as newsletters, that are appropriate to the domain of the CD-ROM can be read by thousands of government researchers. These publications are also read by those who only want to add to their collection of CD-ROMs without the intention of serious application. It is always advised to omit from the advertisement any indication that these discs are available for the asking. Those scientists who have CD-ROM players and have active research in the domain will be interested enough to call or e-mail and the disc can be distributed. It is also wise to determine the scientific intent of the person who requests a CD-ROM. If no serious work is underway and there are no prospects, the request should be denied. This decision is also dependent upon the number of replications that were ordered. With a higher number of replications available for distribution, the policy will, of course, be more lenient.

The OTTER CD-ROM was publicized with an article in both the NASA Information Systems Newsletter and the EOS Observer and with a planned advertisement in the Ecological Society of America Bulletin. Approximately ten requests for the OTTER CD-ROM were received from the article in the first publication, the Information Systems Newsletter.

Section 15   ITEMIZED COSTS

The cost of producing a CD-ROM can vary greatly, depending upon various factors such as amount of data preparation necessary, artwork colors, number of replicates, and the turnaround time. The following are itemized lists of estimated costs for the production of the first OTTER disc, as produced by the PLDS team. There are overall estimates of labor and materials, estimates broken down by type of files, and estimates broken down by production stage.

Labor
      Coordinator, Programmer, Liaison   1 Workyear Total $80K

CD-ROM Costs
      Mastering (1 Disc- 2 color disc artwork) $0.8

Replications (300)	$0.4
Printing (Front Panel)	$0.6
Jewel Box and Insertion	$0.3

Administrative Overhead
Benefit Assessment ($7.9K per WY)	$7.9
Computer Charges ($9K per WY)	$9
Branch and Division Taxes (5%)	$5

| **Total** | **$104K** |

Approximate costs of data preparation are broken down by type of files:

Tabular	$10K
Image	$40K
Documentation	$10K
PDS Label	$10K

Approximate costs are broken down by stage in publication:

Familiarization, Market Analysis, CD-ROM Design	$10K
File Preparation	$50K
File Validation	$20K
Premastering/Mastering (including labor)	$3K
Distribution	$2K

Note: These costs are approximately twice what was predicted at the beginning of the effort, mainly due to the fact that the procedure took twice as long as predicted. It required one person (the programmer) working full time for 6 months to write software to prepare raw and sample data sets and PDS labels for the hundreds of files. The liaison, who worked half time during the six months, interfaced with the investigators, examined images for quality, selected subsets of the raw imagery, and prepared documentation. The coordinator worked one quarter to one half time during this period and performed overall quality assessments.

Section 16   CONCLUSION

This description of the process of CD-ROM production, as compiled during the production of the OTTER CD-ROMs, is intended to educate those who need to publish science data. Many of the steps and thought processes involved in the procedure, such as publication feasibility, must be performed regardless of the technology applied. However, all of the techniques described within this document were applied in its time and will not be applicable in its entirety in the future. Technology changes too fast for any static document such as this. In addition, as data and markets are different in different scientific domains, the procedures involved in data publication will be different.

CD-ROMs of scientific data will be desired by the scientific community for years to come. These carefully-selected discs containing data, documentation, and software will accelerate the progress of research and education in various sciences. A skill that is valuable to many scientists "in the market" is the ability to generate high quality and easy-to-use discs that contain data that are useful for their research and educational purposes.

Section 17   REFERENCES

Data Distribution Laboratory Staff (1993), Producing Multiplatform CD-ROMs, NASA Information Systems Newsletter, Jet Propulsion Laboratory.

Hathaway, Patricia (1992), Generation of Pre-mastered CD-ROM Image Using MakeDisc Software from Young Minds, Inc., Sterling Lessons Learned Report.

Meeson, Blanche, Strebel, Donald, and Landis, David (1992), Producing a CD-ROM: A Workbook, Draft Version 1.0, Internal Document.

Newcomer, Jeffrey and Strebel, Donald (1992), Lossless Compression of Image Data Products on the FIFE CD-ROM Series, Space and Earth Science Data Compression Workshop.

Planetary Data System (1993), Planetary Data System Data Preparation Workbook, Jet Propulsion Laboratory, Internal Document (JPL D-7669, Part 1).

Young Minds, Inc.(1992), Makedisc Version 1.1 User's Manual, Redlands, California.

APPENDICES

OTTER CD-ROM Directory Structure

Directory Tree for the OTTER CD-ROM
Volume 1, Version 1

AAREADME.TXT

AIRPHOTO
    SITE1A
    SITE1OG
    SITE2
    SITE3
    SITE4
    SITE5
    SITE6

AIRSUNP
    ASAS
    SITE1A
    SITE1OG

AVHRR
    Y90M01
    Y90M02
    Y90M03
    Y90M04
    Y90M05
    Y90M06
    Y90M07
    Y90M08
    Y90M09
    Y90M10
    Y90M11
    Y90M12

AVIRIS
    SITE1

CHEM

DAEDALUS
    SITE1
      Y90M03
      Y90M06
      Y90M08
      Y90M10
      Y91M05
    SITE2
      Y90M03
      Y90M06
      Y90M08
      Y90M10

```
        Y91M05
SITE3
     Y90M03
     Y90M06
     Y90M08
     Y90M10
     Y91M05
SITE4
     Y90M03
     Y90M06
     Y90M08
     Y90M10
     Y91M05
SITE5
     Y90M03
     Y90M06
     Y90M08
     Y90M10
     Y91M05
SITE6
     Y90M03
     Y90M06
     Y90M08
     Y90M10
     Y91M05

DOCUMENT

FLDSUNP

METEO

NS001
  SITE1A
     Y90M06
     Y90M08
  SITE1OG
     Y90M06
     Y90M08
  SITE2
     Y90M06
     Y90M08
  SITE3
     Y90M06
     Y90M08
  SITE4
     Y90M06
     Y90M08
  SITE5
     Y90M06
     Y90M08
  SITE6
     Y90M06
     Y90M08
```

SOFTWARE
    MAC
    PC

TIMBER

TIMS
    SITE1A
        Y90M06
        Y90M08
    SITE1OG
        Y90M06
        Y90M08
    SITE2
        Y90M08
    SITE3
        Y90M06
        Y90M08
    SITE4
        Y90M02
        Y90M06
        Y90M08
    SITE5
        Y90M02
        Y90M06
        Y90M08
    SITE6
        Y90M02
        Y90M06
        Y90M08

OTTER AAREADME.TXT File

OTTER: Oregon Transect Ecosystem Research Project
Collected Data, Volume 1, Version 1
Satellite, Aircraft, and Ground Measurements

INTRODUCTION

The Oregon Transect Ecosystem Research (OTTER) Project was a cooperative effort between
NASA and several universities to discern the ecology of western coniferous forests using remote
sensing technology supported by ground observations.

Six Oregon sites across an elevational and climatic gradient were intensively studied. The
transect began at the Pacific coast at the site called Cascade Head, passed through the outskirts of
Corvallis, through a dense Douglas fir forest at Scio, through a mountain hemlock/subalpine fir
community at Santiam Pass, through a Ponderosa pine community near Metolius, and ended at a
site east of Sisters called Juniper. In all, the transect stretched some 300 kilometers west to east.

Goals of the project were to simulate and predict ecosystem processes such as photosynthesis,
transpiration, above-ground production, nitrogen transformation, respiration, decomposition, and
hydrologic processes; combine field, lab, and remote sensing techniques to estimate key

vegetation and environmental parameters; construct a "geo-referenced" database for extrapolation and testing of principles, techniques, and predictions; and verify the predictions through direct measurements of process rates or controls on processes. A further goal of the project was to remotely sense initialization parameters and state variables for a biogeochemical-cycling, forest-stand model.

Measurements were taken using remote sensing instruments aboard satellites (NOAA-11), from instruments flown on NASA's ER-2, DC-8, and C-130 aircraft, from light aircraft, from light experimental (ultralight) aircraft, and on the ground. Field campaigns were coordinated with a Multi-sensor Airborne Campaign (MAC) during 1990 and 1991. The four 1990 data collection periods were timed to coincide with: 1) pre-budbreak at the sites (late March to early April); 2) maximum understory leaf area index (LAI), minimum starch, maximum nitrogen, and maximum LAI in the overstory (late May to early June); 3) maximum LAI and water stress (mid-August); 4) senescence of understory vegetation and reduced LAI and water stress (October). Additional data were gathered in May, 1991. Data are archived at the NASA Ames Research Center.

DISC ORGANIZATION AND CONTENTS

In the top level of this disc, there are a documentation directory, a software directory and several directories with dataset titles. This AAREADME.TXT file contains information about the organization and content of the files on the OTTER CD-ROM and some general use recommendations. A directory entitled DOCUMENT holds general documentation text files. The SOFTWARE directory contains two software packages which can be used to display most OTTER images on this disc. The OTTER data are organized in directories according to dataset.

DOCUMENTATION

The text files in the DOCUMENT directory and their contents are:

	File Name.......Contents
DIRTREE.TXT	Directory structure for this CD-ROM.
FILESTRC.TXT	File structure and naming conventions for all datasets in this CD-ROM.
IMAGEDEX.TXT	Index of the images stored on this CD-ROM, which are sorted by file name, OTTER site, and date and time of acquisition.
KEYWORD.TXT	Keyword assignment statements are described in this file. This is a glossary of various terms from several different types of text files.
RESPLAN.TXT	Original OTTER research plan which gives an overview of the OTTER project.
MACPLAN.TXT	Plan for the OTTER Multi-sensor Airborne Campaign (MAC) as proposed to NASA Headquarters.
MACCAMP.TXT	A description of each of the OTTER MAC data collection periods.
SITELOC.TXT	Names, numbers, and locational coordinates in latitude and longitude for the six OTTER sites and the plots within the sites. Also, the exact location of the sites can be viewed in the digitized aerial photography images stored on this disc using the upper left

and lower right coordinates given in this file.

At the top level of the directory assigned to each of the datasets resides a documentation file with the name (or abbreviated name) of the dataset followed by a .TXT extension. This file contains a description of the instrument or data collection methodology, shows the intended use for the data by the OTTER project, selected references, and other pertinent information useful in applying the data on this CD-ROM. For example, information about OTTER AVHRR data can be found in the file, "AVHRR.TXT", in the "AVHRR" directory.

OTTER DATA

Depending on the dataset, the OTTER data instrument or dataset directories may be divided into sub-directories according to the identification number of the OTTER site and/or the date of acquisition. For medium and low altitude aircraft measurements and for ground measurements, data are often taken separately at plots within the Cascade Head site (site 1) and sometimes at plots within the Scio site (site 3). Following are the data directory names with a brief description of the data within. AIRPHOTO (Digitized Aerial Photographs) One image was taken from the C-130 aircraft at each site in June of 1990 to make up the seven digitized aerial photographic image Files in this directory (an image for each plot within the Cascade Head site is provided). These images can be displayed using the provided software. AIRSUNP (Airborne Sunphotometer). Two tabular files of airborne sunphotometer data taken during two days in August of 1990 from the C-130 aircraft are provided.

ASAS (Advanced Solid State Array Spectrometer) Images for seven tilt angles for two scenes of ASAS data for August of 1990, one for each plot at the Cascade Head site, were taken from the C-130 aircraft and are on this disc. These images are in their native format and cannot be displayed using the provided software.

AVHRR (Advanced Very High Resolution Radiometer) twelve monthly scenes in 1990 over Oregon of geo-registered AVHRR satellite data are provided. These images can be displayed using the provided software.

AVIRIS (Airborne Visible/InfraRed Imaging Spectrometer) One scene of AVIRIS data over the Cascade Head site in August of 1990 taken from the ER-2 aircraft is provided. This image is in its native format and cannot be displayed using the provided software.

CHEM (Canopy Chemistry) One tabular file displays chemistry variables on leaf samples taken at all OTTER sites during all field campaigns. The other tabular file contains summarized monthly specific leaf area data for the Scio site.

DAEDALUS (Daedalus TMS - Thematic Mapper Simulator) Thirty scenes of Daedalus TMS data, consisting of one scene for each of the six sites at the March, June, August, October, 1990, and May 1991 data collection periods are provided. These images taken from the ER-2 aircraft were subsetted from the full flight line and can be displayed using the provided software.

FLDSUNP (Field Sunphotometer) Five tabular files of ground-based field sunphotometer data are provided and consist of one file for each of the field campaigns at all sites.

METEO (Meteorology) Two tabular files of data collected at weather stations near the OTTER sites present hourly and daily measurements of various meteorological variables.

NS001 (NS001 TMS- Thematic Mapper Simulator) 14 scenes of NS001 TMS data, consisting of one scene for each of the sites (two scenes for Cascade Head) during the June and August 1990 data collection periods are provided. These images taken from the C-130 aircraft were subsetted

from the full flight line and can be displayed using the provided software.

TIMBER Five tabular files of timber measurements were compiled from data collected at sites 2, 5, and 6 and at plots within sites 1 and 3.

TIMS (Thermal Infrared Multispectral Scanner) three scenes of TIMS for sites 4,5, and 6 taken from the C-130 aircraft in February of 1990 plus scenes for most sites in June and August of 1990 are provided. These images were subsetted from the full flight line and can be displayed using the software on this CD-ROM.

The ASAS was flown over three of the OTTER sites (1,3, and 5) four times during the data collection periods. These times were 1) high sun, parallel to the plane of the path of the sun; 2) high sun, perpendicular to the plane of the path of the sun; 3) low sun, parallel to the plane of the path of the sun; and 4) low sun, perpendicular to the plane of the path of the sun. The ASAS was flown over the other three OTTER sites (2, 4, and 6) at high and low sun, but only parallel to the plane of the path of the sun.

Other instruments used in OTTER that are flown on the C-130 besides the ASAS are the TIMS and the NS001. Data from those instruments that appear on this CD-ROM were selected to correspond to flights for ASAS flown in the principal plane (parallel to the path of the sun) at high sun. Thus, an image for TIMS for a given site on this CD-ROM was taken on the same heading, at the same altitude, and at the same time as was a corresponding ASAS scene (which includes seven tilt angles).

Because of space limitations on this first compact disc, only two ASAS scenes are included. Subsequent OTTER CD-ROMs will include further ASAS scenes, and these will correspond to the TIMS and NS001 TMS images on this disc.

DISC PREPARATION

The Pilot Land Data System (PLDS) at the Ames Research Center (ARC) prepared the data, documentation, and all supporting files for publication on this CD-ROM. This disc was premastered at the Ames Research Center using Young Minds, Inc. premastering software on a Sun4 Unix machine.

Data files in this CD-ROM follow many of the conventions and structures developed by the Planetary Data System (PDS). For example, each data file is accompanied by a descriptive PDS label file, which in the case of image data, permits easy display of most OTTER images on popular personal computer systems. Documentation and files in alternative formats are provided on the disc for those not accustomed to PDS conventions. Further details on PDS standards and pertinent documentation can be obtained by contacting the PDS operator at (818) 306-6130.

This is the first of several CD-ROMs that will be published containing OTTER project data. Data sets that will appear on subsequent OTTER CD-ROMs in the next year include:
    ASAS
    Airborne SAR (Synthetic Aperture Radar)
    AVIRIS
    Derived Data
    Field Spectrometer Measurements
    Other Field Data
    Forest BGC (Biogeochemical Cycling) Model Data

TECHNICAL INFORMATION

HARDWARE

The OTTER CD-ROM is written in the ISO 9660 format which your CD-ROM device must be able to read. The CD-ROM should be readable by most CD-ROM readers that recognize this format.

Compatibility will be determined by the interaction between the CD-ROM reader, the hardware interface in your computer, and the software interface or driver. The only guarantee that all will work together is to buy them from the same vendor, with a specific requirement that they will work with your host computer (e.g. PC, Macintosh, Sun, VAX).

Throughout the documentation, file names are given in upper case, which is the case in which they appear on the CD-ROM on most systems. However, on some systems, notably Sun workstations, the file names appear in lower case on the CD-ROM. The files can be accessed, contrary to the case-sensitive characteristic of UNIX systems, by entering file names in either upper or lower case.

Recommended CD-ROM Hardware and Drivers
VAX/VMS
Drive: Digital Equipment Corporation (DEC) RRD42, RRD40 or RRD50.
Driver: DEC VMS CD-ROM driver V5.5 and up.

VAX/VMS users: Extended Attribute Records (XARs) are not provided on this disc. All
files   are undefined format files.

VAX/Ultrix
Drive: .   DEC RRD42, RRD40 or RRD50
Driver: Supplied with Ultrix 3.1

Note: Internet users can obtain a copy of the "cdio" software package via anonymous ftp from the "space.mit.edu" server in the file named "src/cdio.shar".

IBM PC
Drive: Toshiba, Hitachi, Sony, NEC or compatible
Driver: Microsoft MSCDEX version 2.2

Note: The latest version of MSCDEX is generally available, usually from CD-ROM hardware vendors. Contact Gary Angelici for assistance in locating a copy.

Apple Macintosh
Drive: Apple CD SC (Sony), NEC or Toshiba
Driver: Apple CD-ROM driver

Note: The Toshiba drive requires a separate driver, which may be obtained from Toshiba.

Sun Micro (SunOS 4.0.x and earlier)
Drive: Delta Microsystems SS-660 (Sony)
Driver: Delta Microsystems driver or SUN sr.o Driver

Note: For questions concerning this driver, contact Denis Down at Delta Microsystems, 415-449-6881.

Sun Micro (SunOS 4.0.x and later)
Drive: Sun Microsystems
Driver: SunOS sr.o driver

SGI (Silicon Graphics, Inc.)
Drive: SGI CD-ROM Drive
Driver: SGI CD-ROM driver

SOFTWARE

Software to display most of the OTTER image data (except ASAS and AVIRIS data) on Macintosh and IBM Personal Computers (and compatibles) is provided on this disc. The popular shareware program, Stuffit, is necessary to extract the execution file for the Macintosh image display program, Image4pds. The public domain software package, Imdisp, is provided for display on IBM compatibles. The label files to be read by the software are provided for all displayable images. The SOFTWARE directory contains the image display programs plus documentation files which aid in the use of the software.

WHOM TO CONTACT FOR INFORMATION

For questions about how to read the OTTER CD-ROM:

Gary Angelici
MS 242-4
Ames Research Center
Moffett Field, CA 94035
(415) 604-5947

Electronic mail addresses:
Internet: gary@pldsa1.arc.nasa.gov
NASAmail: GLANGELICI

ACKNOWLEDGMENTS

The OTTER team cooperatively collected the data presented on this CD-ROM ROM. Acknowledgments are given for each data set, and should be honored when using the data. This CD-ROM (and subsequent CD-ROMs) should be cited as a published volume as Angelici, Skiles, and Popovici (1992) as indicated on the title page of the enclosed booklet.

The support of Dr. Diane Wickland and Dr. Anthony Janetos of NASA Headquarters, Earth Sciences and Applications Division, Ecosystem Dynamics and Biogeochemical Cycles Branch is acknowledged. Also, support of Dr. Maurice Averner of the Biospheric Research Program, Life Sciences Division is acknowledged.

The aircraft flight operations at the NASA/Ames Research Center, the Data Processing Centers at Ames Research Center, Goddard Space Flight Center, and the Jet Propulsion Laboratory are gratefully acknowledged. Gody Spycher of the Long Term Ecological Research (LTER) facility at Oregon State University is thanked for his cooperation in preparing and making OTTER canopy chemistry and meteorology data available for publication.

Jason Hyon, Mike Martin, Sue Hess, Sugi Sorensen, and Marty DeMore of the Planetary Data System (PDS) are acknowledged for providing guidance in the development of this CD-ROM and in providing display software. Matt Ma, system administrator for the PLDS Sun computer, is acknowledged for his system and software configuration work. Thanks go to many others,

including PLDS Project Manager, Dr. Blanche Meeson, who offered valuable advice in the production of this CD-ROM.

Data Management and CD-ROM Production Team: Gary Angelici, Coordinator; Dr. J.W. Skiles; Lidia Popovici.

CD-ROM Artwork: Merin McDonell, David
Faust Photo: Michael Spanner

OTTER File Naming Convention

FILE STRUCTURE AND NAMING CONVENTION
Volume 1, Version 1

FILE STRUCTURE

There are six types of files on this CD-ROM. Their names and functions are as follows:

File Type	Function
Documentation	Text file with documentation about CD-ROM or data
Image data	Binary file of actual image data
Tabular data	Text file of spreadsheet-style data
Label	Text file which describes the image or tabular file
Ancillary	Text file in support of image files
Format	Text file which describes format of the tabular file

DOCUMENTATION FILES

Most documentation files, as described in the AAREADME.TXT file, reside in the DOCUMENT directory, except for dataset description files and software documentation files, which reside in directories for the particular dataset or software execution platform.

IMAGE DATA FILES

Image data files contain actual pixel values that can be displayed using image display software and can be manipulated through most image processing programs. Image data files are provided for several remote sensing instruments: AVHRR, AVIRIS, ASAS, Daedalus TMS, NS001 TMS, and TIMS. Image data of aerial photography that has been digitized is also provided on this CD-ROM.

The records of an image file on this CD-ROM are all the same length, and there is no embedded information to indicate the beginning or end of a record. These fixed-length records allow any part of a file to be accessed directly without the need to pass through the file sequentially. Image files contain binary data with no carriage-control information in them.

Individual bands for all images (except for the AVIRIS and ASAS scenes, which are stored in their native formats) are provided in separate files with no headers of any kind. For example, there are 8 files of imagery for each NS001 flight line, one for each band.

For the three datasets of Daedalus TMS, NS001 TMS, and TIMS, the images on this CD-ROM have been extracted from the full flight line to create subset images which cover only the area surrounding the OTTER sites.

TABULAR DATA FILES

OTTER tabular data files of field sunphotometer, airborne sunphotometer, chemistry, meteorology, and timber measurements contain data collected or generated by OTTER investigators. These data files are in the format of a table with data measuring several variables at different time intervals or data points.

Tabular files are formatted so that they may be read directly into many database management systems and spreadsheet programs on various computers. All fields are separated by commas and character fields are enclosed in double quotation marks ("). The records are of fixed length, and the last two bytes of each record contain the ASCII carriage return and line feed characters. This allows a tabular file to be treated as a fixed length record file on computers that support this file type and as a normal text file on other computers.

LABEL FILES

For each image data or tabular data file on the OTTER CD-ROM, there is a label file. This PDS file contains descriptive information about the image or tabular file. For image data from the Daedalus TMS, NS001 TMS, TIMS, digitized aerial photography, and AVHRR datasets, these files can be used to easily display the associated imagery files using the provided display software without the need for the user to enter descriptive information. Documentation about displaying the imagery is offered in text files within the SOFTWARE sub-directory.

For tabular data files, the format of the data tables is provided in the accompanying label file along with the descriptive information. Format files also provide the same basic format information in a condensed style.

At the end of this file, a description of label files plus an example of a label file are provided.

ANCILLARY FILES

Files that provide ancillary information about the main image file include calibration files that accompany the AVIRIS instrument data. With the image files of Daedalus TMS, NS001 TMS and TIMS, two sets of ancillary files have been generated and are provided on this CD-ROM. First, files of housekeeping information for each band of aircraft data, offering values for such variables as black body temperature for each aircraft scan line, are provided in a tabular format and can be processed using common personal computer spreadsheet or database programs. A format file is provided for this file. Second, a summary file of calibration and other ancillary information for each subset scene contains means and standard deviations for various values in the housekeeping data along with general geographic and flight geometry values.

FORMAT FILES

For each dataset that has data in a tabular form, such as field sunphotometer, there is a file which lists the fields, their column placement in the file, and other descriptive information. The same basic information in the PDS style is provided in the label file.

Format files are also used in describing the format of other data files. Selected image datasets have header records which are described using a format file. For example, the format of the ASAS image header record is provided in a format file. Also, the formats of the housekeeping files for Daedalus TMS, NS001 TMS and TIMS data are described format files.

DOCUMENT, LABEL, ANCILLARY, and FORMAT FILES. All document, label, ancillary, and format files contain 80-byte fixed-length records, with a carriage return character (ASCII 13)

in the 79th byte and a line feed character (ASCII 10) in the 80th byte. This permits the files to be read by the MacOS, DOS, Unix and VMS operating systems and printed or displayed on a terminal.

FILE NAMING CONVENTION

The file naming convention varies slightly between the datasets. All files (except for documentation files) follow the same rules regarding the first two characters of the file name and the extension (the portion to the right of the period). The characters after the first two and before the period vary in format depending upon the dataset. The following displays the structure of each file name:

AABBBBBB.EXT where AA are the First Two Characters, BBBBBB are the Intermediate Characters, and EXT is the Extension

While the next sections describe the general file naming convention, each dataset description file (e.g. DAEDALUS.TXT) in the dataset directories describe in detail the file name structure for the files associated with that dataset.

Note: To preserve unique file names throughout the disc while observing the 8 character file name limit, this approach was used to name data files. While the file names themselves do not typically provide sufficient information to determine site location or date of acquisition, this information can be found for each image file in the IMAGEDEX.TXT file. The names of the files do, however, identify the dataset and whether the files are data files or are supporting text files. The file names also permit the identification of the files which contain individual bands of image data.

FIRST TWO CHARACTERS

The first two characters of each OTTER file name is a code for the dataset. The coding scheme is as follows:

Dataset Code = Dataset

AH = AVHRR
AP = Airborne Sunphotometer
AS = ASAS
AV = AVIRIS
CH = Canopy Chemistry
DA = DAEDALUS TMS
DP = Digitized Air Photographs
FS = Field Sunphotometer
ME = Meteorology
NS = NS001 TMS
TB = Timber Measurements
TI = TIMS

EXTENSION

According to PDS rules, the following extensions correspond to the types of data within the file. The extensions and their definitions are as follows:

Extension____Type of Data

.TXT   Documentation file
.IMG   Image file
.TAB   Tabular data file
.LBL   Label file
.DAT   Ancillary file
.FMT   Format File

INTERMEDIATE CHARACTERS

The intermediate characters in the file name vary between datasets. Examples of file names for the data and accompanying label, ancillary and format files for each image and tabular dataset are provided.

Image Data Files

Depending upon the type, image data is stored on this CD-ROM in one of two band storage formats: separate files per band or native format. The image data for Daedalus TMS, NS001 TMS, TIMS, digitized aerial photography, and AVHRR are stored in separate files per band, and AVIRIS and ASAS data are stored in their native format. The file naming convention differs slightly for each format.

Separate Files Per Band

For images stored in separate files per band, file names consist of the two letter code for the instrument that acquired the data (as described above), the last three digits of the PLDS/Ames on-line database entry identification number, the letter "B" for band, and a 2 digit number signifying the band number of the file.

An example file name for the first band of a Daedalus TMS image data file with the last three digits of the entry identifier of "037" would be:

DA037B01.IMG

There is also a PDS label file for each band, with an identical file name, except that the extension is .LBL. This file can be used to display the .IMG file in the software provided on this CD-ROM.

Data sets that have ancillary files with housekeeping and summary information are Daedalus TMS, NS001, and TIMS. The file name for the file containing the housekeeping data for the first band of the above Daedalus TMS image follows the convention of the image file would be DA037B01.DAT. Also, the formats of the housekeeping (.DAT) files for the Daedalus TMS, NS001, and TIMS datasets are described in the .FMT files entitled DAEDHSK.FMT, NS001HSK.FMT, and TIMSHSK.FMT, respectively.

Because the summary file, which summarizes the information from the housekeeping files applies to an entire flight line (and not by band), the convention is to remove the band designation in the file name and insert "HED", as in DA037HED.DAT.

The file names for the NS001 and TIMS data, including the housekeeping and header files, are constructed in an identical fashion. The file names for the AVHRR data and digitized aerial photographs are also constructed in a similar fashion, except that there are no associated housekeeping or summary files. Therefore, there are no .DAT or .FMT files in the directories for those datasets.

Native Format For AVIRIS images, which consist of several files of imagery and calibration data, the file name is composed of the two letter code for the instrument that acquired the data (AV), the last three digits of the PLDS on-line database entry identification number, and a letter signifying the type of file. The letter codes and the AVIRIS file are:

Code	Type of File
A	Image file
B	Navigation calibration file
C	Engineering calibration file
D	Dark Current calibration file
E	Offset calibration file
F	Dropout calibration file
G	Vignetting calibration file
H	Radiometric calibration file
I	Spectral response (fwhm) calibration file
J	Spectral wavelength calibration file
K	Laboratory on-board calibration file
L	Noise equivalent radiance file
M	Noise spike replace list
N	Pre-calibration file

For example, the file names for the AVIRIS files with the last three digits of the entry identifier of "033" would be:

Image file = AV033A.IMG
Navigation calibration file = AV033B.DAT
Engineering calibration file = AV033C.DAT
Dark Current calibration file = AV033D.DAT
Offset calibration file = AV033E.DAT
Dropout calibration file = AV033F.DAT
Vignetting calibration file = AV033G.DAT
Radiometric calibration file = AV033H.IMG
Spectral response (fwhm) calibration file = AV033I.DAT
Spectral wavelength calibration file = AV033J.DAT
Laboratory on-board calibration file = AV033K.DAT
Noise equivalent radiance file = AV033L.DAT
Noise spike replace list = AV033M.DAT
Pre-calibration file = AV033N.DAT

For ASAS images, which consist of seven images which each view the same scene from different tilt angles, the file name is composed of the two letter code for the instrument (AS), the last three digits of the PLDS on-line database entry identification number followed by three characters which designate the tilt angle of the instrument for that image. The code is as follows:

Code	Tilt Angle
Fdd	Forward tilt angle of "dd" degrees
Bdd	Backward tilt angle of "dd" degrees
NAD	Nadir view

For example, ASAS data file with the name, AS001F45.IMG, would contain image data for the 45 degree forward tilt angle image for the scene represented by the entry identifier with 001 as its last three characters. There are also typically six other files (corresponding to the other six tilt

47

angles) for the scene:

    AS001F30.IMG
    AS001F15.IMG
    AS001NAD.IMG
    AS001B15.IMG
    AS001B30.IMG
    AS001B45.IMG

The PDS label files have the same file name, except that the extension is .LBL. Also, the format of the ASAS image header record, in PDS format, can be found in the file, ASAS.FMT.

Tabular Data Files

Due to the nature of the data, tabular data vary in several ways and defy a consistent file naming structure, except that all carry the .TAB extension. Besides the two initial characters, the intermediate characters are different for each dataset.

For Field Sunphotometer data, the six characters after the two character data set code display the year and month of data collection. The format is: YyyMmm where yy is the number of the year, and mm is the number of the month.

For example, tabular data collected in August of 1990 is stored in the file, FSY90M08.TAB, the PDS label for the data is stored in the file, FSY90M08.LBL, and the format for the tabular data file is stored in the file, FLDSUNP.FMT.

For Airborne Sunphotometer data, the six intermediate characters consist of the month and the day of data acquisition. The format is: MmmDdd where mm is the number of the month, and dd is the number of the day. For example, data for August 13, 1990 collection day would be stored in the file, APM08D13.TAB and the PDS label is stored in the file, APM08D13.LBL. The format for the data is stored in all airborne sunphotometer tabular data files is found in the file, AIRSUNP.FMT.

For Canopy Chemistry data, the five characters after the initial two characters designate whether the data are daily or monthly data. The format is:

CHxxxxx where xxxxx is either DAILY or MONTH.

For example, the monthly chemistry data are stored in the file, CHMONTH.TAB and the PDS label is stored in the file, CHMONTH.LBL. Because the format is different between the daily and monthly files, two format files are provided. The format for the data in the example is stored in the file, CHMONTH.FMT.

For Meteorology data, the six characters after the initial two characters designate whether the data are hourly or daily data. The format is: MExxxxxx where xxxxxx is either HOURLY or DAILY. For example, the hourly meteorology data is stored in the file, MEHOURLY.TAB and the PDS label is stored in the file, MEHOURLY.LBL. Because the format is different between the hourly and daily files, two format files are provided. The format for the data in the example is stored in the file, MEHOURLY.FMT.

For Timber data, the intermediate characters after the initial two characters designate the site at which the data were collected. The format is: TBSTExxx where xxx is one, two or three characters for the OTTER site number and plot designator.

For example, the timber data collected at the Old Growth (OG) plot of the Cascade Head site (site 1) is stored in the file, TBSTE1OG.TAB and the PDS label is stored in the file, TBSTE1OG.LBL. The format for the data in all timber tabular data files is found in the file, TIMBER.FMT.

PDS LABEL DESCRIPTION

The data sets on this CD-ROM are accompanied by detached PDS labels to describe and point to data files. The PDS label contains descriptive information about the data file and objects within the file. The label consists of keyword statements that conform to the Object Description Language (ODL) developed by the PDS project. There are three types of ODL statements within a label: structural statements, keyword assignment statements, and pointer statements.

Structural statements provide a shell around keyword assignment statements to delineate which data object the assignment statements are describing. The structural statements are:

1) OBJECT = object_name
2) END_OBJECT
3) END

The OBJECT statement begins the description of a particular data object and the END_OBJECT statement signals the end of the object's description. All keyword assignment statements between an OBJECT and its corresponding END_OBJECT statement describe the particular object named in the OBJECT statement. The END statement terminates a label. It must appear as a single statement that contains only the word END.

A keyword assignment statement contains the name of an attribute and the value of that attribute. These statements have the following format:

name = value

Values of keyword assignment statements can be numeric values, literals, or text strings.

Pointer statements are a special class of keyword assignment statements. These pointers are expressed in the ODL using the following notation:

^object_name = location

If the object is in the same file as the label, the location of the object is given as an integer representing the starting record number of the object, measured from the beginning of the file. The first label record in a file is record 1. Pointers are useful for describing the location of individual components of a data object. Pointer statements are also used for pointing to data or information stored in separate files. An example of a pointer to information stored in a separate file is shown below:

^DESCRIPTION = "logical_file_name"

The value of "logical_file_name" is the name of the file containing the description. An example of a pointer to a record within a file is:

^IMAGE = ("logical_file_name",2)

This indicates that image data begins at record 2 in file "logical_file_name".

Each keyword assignment is stored as a single variable-length record. An example of a label is shown below. The file, KEYWORD,TXT, provides a detailed description of each keyword found in the labels.

EXAMPLE OF A DETACHED PDS IMAGE LABEL

CCSD3ZF0000100000001NJPL3IF0PDS200000001 = SFDU_LABEL

/* PDS label for an OTTER Daedalus TMS image */

RECORD_TYPE.= FIXED_LENGTH
RECORD_BYTES = 716
FILE_RECORDS = 879

/* Pointers to objects */

^IMAGE = ("DA269B03.IMG", 1)

/* Image description */

DATA_SET_ID = 'DAEDALUS TMS'
PRODUCT_ID = "DA269B03"
INVESTIGATOR = "MICHAEL SPANNER"
INSTRUMENT_NAME = 'DAEDALUS THEMATIC MAPPER SIMULATOR'
START_DATE_GMT = 1990-10-19T19:58:44Z
STOP_DATE_GMT = 1990-10-19T20:00:09Z
PROJECT = "OTTER"
GEOGRAPHICAL_REGION = "OREGON TRANSECT"
FULL_FLIGHT_NUMBER = "91-019"
FLIGHT_NUMBER = 19
FLIGHT_PROJECT_NUMBER = "90L223D"
CHECK_POINTS = "D-C"
LATITUDE_CENTER = 45.0497
LONGITUDE_CENTER = -123.9667
LATITUDE_START = 45.2128
LONGITUDE_START = -123.9412
LATITUDE_STOP = 44.8865
LONGITUDE_STOP = -123.9921
SITE_NAME = "CASCADE HEAD"
SITE_NUMBER = 1
PLATFORM = "ER-2"
ALTITUDE_MSL = 19817
NADIR_PIXEL_SIZE = 25.4
CERTIFICATION_LEVEL = "PI CHECKED"
CERTIFICATION_COMMENTS = "CLEAR"
BAND_NUMBER = 3
BAND_SPECTRAL_RANGE = "630 - 690 nm"
^DESCRIPTION = "DAEDALUS.TXT"

/* Description of objects */
OBJECT = IMAGE
LINES = 879
LINE_SAMPLES = 716
SAMPLE_TYPE = BYTE

```
SAMPLE_BITS = 8
BANDS = 1
END_OBJECT = IMAGE
END
```

Example OTTER Instrument Documentation

Thematic Mapper Simulators (TMS)
OTTER Data Description

NS001 Data Files on This CD-ROM: Within the NS001 directory, there are seven sub-directories (one for each OTTER site and two for the Cascade Head site), each containing two sub-directories (one for each of the OTTER aircraft data collection periods of June 1990 and August 1990). Within each of these sub-directories, there is one scene of NS001 TMS data which is a subset of the full flight line and includes the immediate area surrounding the OTTER site. The individual bands of the NS001 TMS have been placed in separate files.

For example, the file name for band 3 of the NS001 TMS data for the August 1990 data for the Alder Plot of the Cascade Head site 1 is NS227B03.IMG. This file and its associated files are located in sub-directory, NS001/SITE1A/Y90M08, where 90 is the Year 1990 and 8 is the August.

In addition to the image (.IMG) files, there are files of housekeeping information (.DAT) for each band of data which contain tabular data for each scan line. For example, the file name for the housekeeping file for the band mentioned above is NS227B03.DAT. The format for all of the NS001 TMS housekeeping files is described in the file, NS001HSK.FMT, located at the top level of the NS001 directory.

A summary file of calibration and other ancillary information for each scene is also provided. The file name for this file has a .DAT extension and the characters "HED" before the period. For example, the summary file for the scene mentioned above would be stored in NS227HED.DAT.

The files in the directory with the same file names with a .LBL extension contain descriptive information about the image file.

The image files can be displayed using the display software provided on this CD-ROM.

Instrument Description:

The Thematic Mapper Simulator (TMS) instruments are designed to simulate spectral, spatial, and radiometric characteristics of the Thematic Mapper sensor on the Landsat-4 and 5 spacecraft. The two instruments used in OTTER are similar, but they are flown at different altitudes thereby yielding data with different resolutions. The NS001 TMS is generally flown at medium altitudes and provides 12.2 meter resolution at nadir at an altitude of 16,000 feet (4,864 meters). The Daedalus TMS is flown at higher altitudes and provides a ground resolution of 25 meters at an altitude of 65,000 feet (19,760 meters). The NS001 TMS is flown aboard NASA's C-130 aircraft based at the NASA Ames Research Center. The Daedalus TMS is flown aboard the NASA ER-2 aircraft.

Although the TMS sensors are very similar, they differ slightly from each other and from the Landsat TM instruments. Both TMS instruments have 7 spectral channels that are very similar to those of the TM sensor. However, they both have additional channels; the NS001 TMS has one additional IR channel and the Daedalus TMS has five additional channels.

The Daedalus TMS spectral bands are as follows:

Daedalus	Channel	TM Band	Wavelength, um
1	A	0 42-0 45	
2	1	0 45-0 52	
3	2	0 52-0 60	
4	B	0 60-0 62	
5	3	0 63-0 69	
6	C	0 69-0 75	
7	4	0 76-0 90	
8	D	0 91-1 05	
5		1 55-1 75	
10	7	2 08-2 35	
11	6	8 5-14 0 low gain	
12	6	8 5-14 0 high gain	

Other parameters for the Daedalus TMS are:

IFOV 1.25 mrad	
Ground Resolution	81 feet (25 meters) at 65,000 feet
Total Scan Angle	43 degrees
Swath Width	8.4 nmi (15.6km) at 65,000 feet
Pixel/Scan Line	716
Scan Rate	12.5 scans/second
Ground Speed	400 kts (206 m/sec)

The NS001 TMS spectral bands are as follows:

NS001	Channel	TM Band	Wavelength, um
1	1	0.45-0.52	
2	2	0.52-0.60	
3	3	0.63-0.69	
4	4	0.76-0.90	
5		1.00-1.30	
6	5	1.55-1.75	
7		2.08-2.35	
8	6	10.40-12.5	

Other parameters for the NS001 TMS are:

IFOV	2.5mrad
Total Scan Angle	100 degrees
Pixels/Scan Line	700

OTTER Use of TMS:

TMS data are to be used to determine leaf area index of the OTTER sites as well as various vegetation indices (e.g. band 4 divided by band 3).

OTTER Data Acquisitions (GMT):

Daedalus TMS
    3-15-88
    4-1-88

```
5-20-88
6-1-89
2-22-90
3-21-90
6-25-90
6-26-90
8-13-90
8-14-90
10-19-90
10-24-90
5-22-91
```

NS001 TMS
```
6-19-90
6-20-90
6-21-90
8-13-90
8-14-90
8-15-90
```

Note:
The October 19, 1990 image has a problem with the black body reference temps for all bands; the value given is 7000, which is clearly in error (all the other scenes had temperatures around 1200). (as per Joe Glassy)

Data Decommutation:
OTTER contact   Mike Spanner
TGS Technology
Mail Stop 242-4
NASA/Ames Research Center
Moffett Field, California 94035
415/604-3620

References:
Peterson, D. L., Westman, W. E., Stephenson, N. J., Ambrosia, V. G., Brass, J. A., and Spanner, M. A. 1986 Analysis of forest structure using Thematic Mapper Simulator data. IEEE Transactions on Geoscience and Remote Sensing GE-24(1):113-121.

Peterson, D. L., Spanner, M. A., Running, S. W., and Teuber, K. B. 1987. Relationship of Thematic Mapper Simulator data to leaf area index of temperature coniferous forests. Remote Sensing of Environment 22:323-341.

Spanner, M. A., Peterson, D. L., Hall, M. J., Wrigley, R. C., Card, D. H., and Running, S. W. 1984. Atmospheric effects on the remote sensing estimation of forest leaf area index. Eighteenth International Symposium on Remote Sensing of Environment, Paris, France. Proceedings published by Environmental Research Institute of Michigan.

Author and date of extract: Jay Skiles March 7, 1991

Example OTTER PDS Label File

```
/* PDS label for an OTTER NS001 TMS image */
RECORD_TYPE                    = FIXED_LENGTH
```

```
RECORD_BYTES                = 699
FILE_RECORDS                = 512
/* Pointers to objects */
^IMAGE                      = ("NS291B01.IMG",1)
/* Image description */
DATA_SET_ID                 = 'NS001_TMS'
PRODUCT_ID                  = "NS291B01"
INVESTIGATOR                = "MICHAEL SPANNER"
INSTRUMENT_NAME       =
'NS001 THEMATIC MAPPER SIMULATOR'
START_DATE_GMT              = 1990-06-19T19:38:25
STOP_DATE_GMT               = 1990-06-19T19:41:12
POLITICAL_SUBDIVISION       = "OREGON"
PROJECT                     = "OTTER"
FULL_FLIGHT_NUMBER          = "90-003-03"
FLIGHT_NUMBER               = 3
LINE_NUMBER                 = 1
RUN_NUMBER                  = 2
FLIGHT_PROJECT_NUMBER       = "90L314D"
LATITUDE_CENTER             = 44.0288
LONGITUDE_CENTER            = -121.2969971
LATITUDE_START              = 44.0288
LONGITUDE_START             = -121.2969971
LATITUDE_STOP               = 44.0288
LONGITUDE_STOP                  = -121.2969971
SITE_NAME                   = "WARINGS WOODS"
SITE_NUMBER                 = 2
SCAN_LINES                  = 512
PLATFORM                    = "C-130"
ALTITUDE_AGL                = 4572
NADIR_PIXEL_SIZE            = 39.3
CERTIFICATION_LEVEL         = "PI CHECKED"
CERTIFICATION_COMMENTS      = "GOOD RUN"
BAND_NUMBER                 = 1
BAND_SPECTRAL_RANGE         = "8.2 - 8.6"
^DESCRIPTION                = "NS001.TXT"
/* Description of objects */
OBJECT                      = IMAGE
LINES                       = 512
LINE_SAMPLES                = 699
SAMPLE_TYPE                 = BYTE
SAMPLE_BITS                 = 8
BANDS                       = 1
END_OBJECT                  = IMAGE
END
```

Image Display Test Plan
For AVHRR, Daedalus TMS, Digitized Aerial Photo, NS001, and TIMS Data

<u>Processing</u>

1) Select two flight lines or scenes from different dates

2) Execute conv_images to create separate image files per band without headers

For Daedalus TMS, NS001, TIMS, you will need:
- FSRs for the flight lines (plus spreadsheets for the C-130 flights)
- Radiances and offsets- from Spanner list (Daedalus: Enter 99.9 for the last channel)
  or aircraft program list (NS001: Enter from "Calibration Factor")
- store on disk all image, housekeeping and header files

Image Verification

Send each image band file to Mac and display using the "Verifying" procedure attached
  (is there enough disk space?)
- check that the correct band is there
- check that all four corners of subset are correct

Housekeeping and Summary File Verification (for Daedalus TMS, NS001, and TIMS only

Print the first few pages of the housekeeping file for each band of each flight line/run

Print the housekeeping file format document

- check housekeeping values by running SCANHEAD on VAX

Print the entire summary file for each flight line/run
- check that values for items in summary file approximate the housekeeping data
- hand calculate values (such as pixel size) where possible to check summary values
- give all documentation to investigator (make sure it is all ready to go) to verify- if possible,
have him look at screen at the same time as looking at housekeeping and summary files
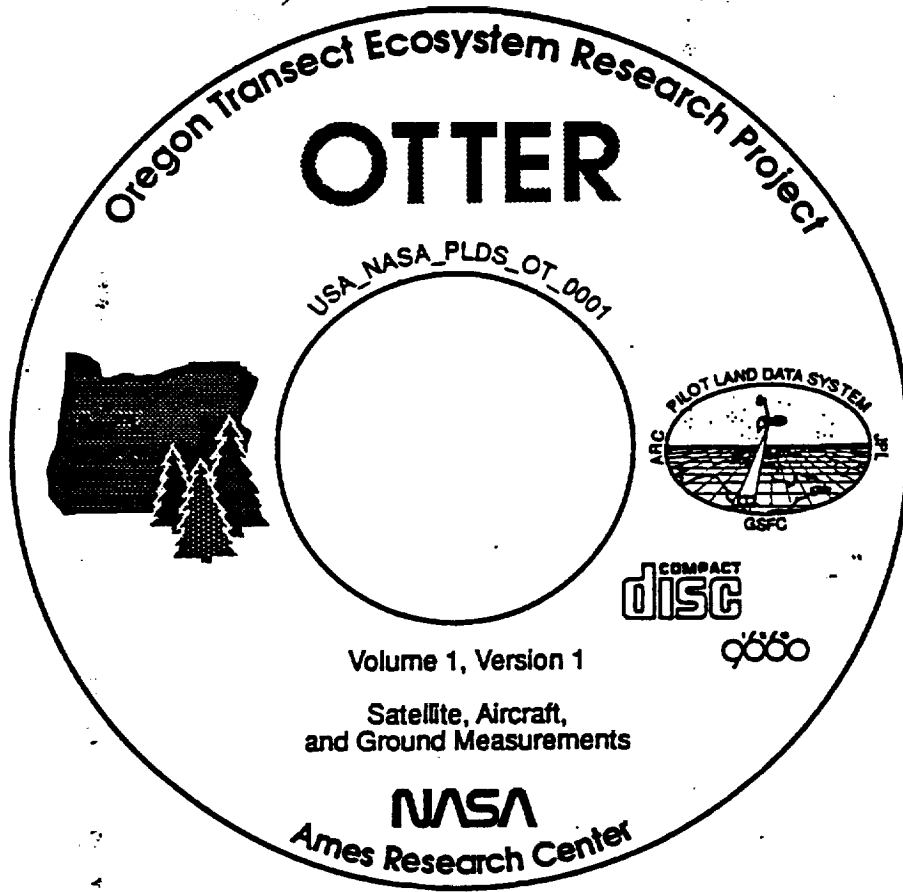
Purchase Request for First OTTER Disc

STERLING SOFTWARE		

Sterling Federal Systems Group
NASA Ames Division

PROCUREMENT REQUEST

DATE INITIATED: May 18, 1992
PR NO.: (FOR OFFICE USE ONLY) JJ1 6

STERLING REQUESTOR: (PLEASE PRINT) Gary Angelici	ORIG. DEPT. OR PRIME CONTRACT: NAS2-13210 __X__ NAS2-12311 ____ NAS2-13403 ____ PAO ____ OTHER: (SPECIFY)	TASK: 402	SUBTASK:
PHONE: 4-5947		CALL FOR PICK UP: Gary Angelici	PHONE 4-5947

ADP SYSTEM: __SGE__ ADP SUBSYSTEM: __SG GRAPHIC__ FOR ADP INFO CALL ADP MGMT OFFICE X4-4012, X4-4798 OR X4-0217	SUBCONTRACT YES: _____ NO: X _____ IF YES: New Subcontract _____ Existing Subcontract _____ Original PR No. _____

TECHNICAL CONTACT: Gary Angelici	PHONE: 4-5947	PERIOD OF PERFORMANCE: (IF APPLICABLE) N/A

JUSTIFICATION:
    CD-ROM Mastering of pre-mastered magnetic tapes are required for Sterling to produce CD-ROMs of OTTER data.

RECOMMENDED SOURCES (IF ANY) Disc Manufacturing, Inc.. Nimbus Information Systems	PHONE: (800) 433-DISC (Pam Krieger of Disc Manufacturing, Inc.) (804) 985-1100 (Nimbus Info Systems)	SOLE OR SINGLE SOURCE (If yes, attach vendor recommendation memo) NO X _____ YES _____

This is advanced acknowledgement only and does not constitute Advanced Notification or Prior Consent of the Contracting Officer as required by FAR 52.244-2.
NASA Task Requestor _____ Org 'SGE Date 5/18/92 NASA COTR _____ Date 5/26/92

ITEM NO.	QTY	DESCRIPTION	UNIT PRICE	ESTIMATED TOTAL
1	2	Proof Discs of OTTER data in ISO 9660 Format (5 day Turnaround)- 1st Draft	$150.00	$300.00
2	1	Proof Disc in ISO 9660 Format (5 day Turnaround)- 2nd Draft (Optional)	$150.00	$150.00
3	1	Mastering of OTTER Data in ISO 9660 Format (3 day Turnaround)	$950.00	$950.00
		- all mastering (proofs and masters) must accept Makadisc format on 8mm Tape		
4	300	Replications of Discs generated in item 3 (with 2 color disc label)	$1.30	$390.00
		- Spin Coat lacquer, Track pitch no less than 1.6 microns, Block Error Rate (BLER) no		
		greater than 50, error averaged over 10-second intervals. No peak BLER greater		
		than 100 anywhere on disk. All other physical measurements must be safely within		
		Sony-Philips Yellow Book Standard. Documentation of adherence to above required..		
5	300	Jewel Box, Disk Insertion and Shrink Wrap for Item 4 *(Booklet Insertion)*	$.35	$105.00
		Transportation Charges (if applicable)		

SPU

REQUIRED DELIVERY DATE: $October 31, 1992	MAXIMUM AUTHORIZED EXPENDITURE: $2500.00		TOTAL	$1895.00

Disc-face artwork final design. The outline of Oregon, the trees, and the word "OTTER" are rendered in Pantone #370, Forest Green. All other marks are rendered in black.

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE November 1993	3. REPORT TYPE AND DATES COVERED Contractor Report

4. TITLE AND SUBTITLE

Publication of Science Data on CD-ROM: A Guide and Example

5. FUNDING NUMBERS

NAS2-13210

6. AUTHOR(S)

Gary Angelici* and J. W. Skiles**

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

*Sterling Software, Inc.
1121 San Antonio Road
Palo Alto, CA 94303

**Johnson Controls World Services
7315 N. Atlantic Ave.
Cape Canaveral, FL 32920

8. PERFORMING ORGANIZATION REPORT NUMBER

A-94024

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

National Aeronautics and Space Administration
Washington, DC 20546-0001

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

NASA CR-4560

11. SUPPLEMENTARY NOTES

Point of Contact: Gary Angelici, Ames Research Center, MS 242-B, Moffett Field, CA 94035-1000
(415) 604-5947

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Unclassified-Unlimited
Subject Category – 43

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

CD-ROM (Compact Disk-Read Only Memory) is becoming the standard media not only in audio recording, but also in the publication of data and information accessible on many computer platforms. Little has been written about the complicated process involved in creating easy-to-use, high quality, and useful CD-ROMs containing scientific data. This document is a manual designed to aid those who are responsible for the publication of scientific data on CD-ROM. All aspects and steps of the procedure are covered, from feasibility assessment through disk design, data preparation, disc mastering, and CD-ROM distribution. General advice and actual examples are based on lessons learned from the publication of scientific data for an interdisciplinary field experiment. Appendices include actual files from a CD-ROM, a purchase request for CD-ROM mastering services, and the disk art for the first disk published for the project.

14. SUBJECT TERMS

CD-ROM, Data publication, Manual

15. NUMBER OF PAGES

64

16. PRICE CODE

A04

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified		