British Geological Survey

NATURAL ENVIRONMENT RESEARCH COUNCIL

QC procedures for archiving seismic data to the National Hydrocarbons Data Archive (NHDA)

Information Management Programme

Internal Report IR/06/121

BRITISH GEOLOGICAL SURVEY

INFORMATION MANAGEMENT PROGRAMME

INTERNAL REPORT IR/06/121

QC procedures for archiving seismic data to the National Hydrocarbons Data Archive (NHDA)

J. Bulat

Contributor/editor

Colin Graham

.

Keywords

Seismic, QC, NHDA

Bibliographical reference

BULAT, J. QC procedures for archiving seismic data to the National Hydrocarbons Data Archive (NHDA). *British Geological Survey Internal Report*, IR/06/121. 42pp.

Edinburgh, Scotland British Geological Survey 2006

BRITISH GEOLOGICAL SURVEY

The full range of Survey publications is available from the BGS Sales Desks at Nottingham, Edinburgh and London; see contact details below or shop online at www.geologyshop.com

The London Information Office also maintains a reference collection of BGS publications including maps for consultation.

The Survey publishes an annual catalogue of its maps and other publications; this catalogue is available from any of the BGS Sales Desks.

The British Geological Survey carries out the geological survey of Great Britain and Northern Ireland (the latter as an agency service for the government of Northern Ireland), and of the surrounding continental shelf, as well as its basic research projects. It also undertakes programmes of British technical aid in geology in developing countries as arranged by the Department for International Development and other agencies.

The British Geological Survey is a component body of the Natural Environment Research Council.

British Geological Survey offices

Keyworth, Nottingham NG12 5GG

☎ 0115-936 3241 Fax 0115-936 3488
e-mail: sales@bgs.ac.uk
www.bgs.ac.uk
Shop online at: www.geologyshop.com

Murchison House, West Mains Road, Edinburgh EH9 3LA

☎ 0131-667 1000 Fax 0131-668 2683
e-mail: scotsales@bgs.ac.uk

London Information Office at the Natural History Museum (Earth Galleries), Exhibition Road, South Kensington, London SW7 2DE

☎ 020-7589 4090 Fax 020-7584 8270
☎ 020-7942 5344/45 email: bgslondon@bgs.ac.uk

Forde House, Park Five Business Centre, Harrier Way, Sowton, Exeter, Devon EX2 7HU

☎ 01392-445271 Fax 01392-445371

Geological Survey of Northern Ireland, Colby House, Stranmillis Court, Belfast BT9 5BF

☎ 028-9038 8462 Fax 028-9038 8461

Maclean Building, Crowmarsh Gifford, Wallingford, Oxfordshire OX10 8BB

☎ 01491-838800 Fax 01491-692345

Columbus House, Greenmeadow Springs, Tongwynlais, Cardiff, CF15 7NE

☎ 029–2052 1962 Fax 029–2052 1963

Parent Body

Natural Environment Research Council, Polaris House, North Star Avenue, Swindon, Wiltshire SN2 1EU

☎ 01793-411500 Fax 01793-411501
www.nerc.ac.uk

Contents

APPENDICES

FIGURES

TABLES

Summary

This report describes the QC procedures for the archival of seismic data to the National Hydrocarbons Data Archive (NHDA). The NHDA has been set up with DTI agreement as a means of allowing companies to waive their statutory requirement to hold and maintain all exploration data in perpetuity by donating re-mastered seismic data sets to the NHDA along with a single ‘dowry’ payment to BGS to cover archival costs. The NHDA could then provide third parties these data for the cost of reproduction.

Three principle data types are archived; original raw data i.e. field data usually provided on SEGD tapes, final and near-final processed data provided on SEGY tapes and navigation data provided in UKOOA format. Other data in addition are lodged with the NHDA such as acquisition and processing reports. Each data type requires different QC processes. However, the overall aims are the same for each i.e. to ensure that all the data that was agreed to be archived has been received in full and is fully internally consistent.

In addition to describing the QC procedures, this manual lists the various UNIX scripts that have been created to perform the QC tasks to act as a permanent record and reference, so that they can be recreated easily in the event of disc malfunction.

1 Introduction

This manual describes the procedures needed for the archival of seismic data sets donated to the National Hydrocarbons Data Archive (NHDA). The NHDA has been set up with DTI agreement as a means of allowing companies to waive their statutory requirement to hold and maintain all data acquired as part of a DTI Licence in perpetuity. This is achieved by the Licensee transferring an agreed set of data, acquired under the Licence Agreement, to the NHDA, along with a dowry payment. BGS then takes on the twin obligations of managing the data in perpetuity and making it available to the public at low cost. Seismic data are included in this arrangement. Both 2D and 3D seismic data sets are archived, and these are supplied as four standard data types:

- Raw field data (as SEGD files),
- Processed data volumes (as SEGY files),
- Navigation files (typically UKOOA P1/90 but also other UKOOA varients)
- Ancillary documents such as observer's logs, acquisition and processing reports.

The SEGY & SEGD files will be provided on either IBM3590 or LTO2 cartridges. Other data types may be supplied on CD or DVDs as well as IBM3590. In exceptional circumstances, other types of media will be accepted (for example, IBM 3480) where the seismic data are the output of special seismic processing (e.g. PSDM and PSTM data). This is in addition to the 'standard' raw field and processed data volumes, which must be supplied on IBM3590 or LTO2. The seismic data on non-standard media are indexed and archived only – the data are not transferred to other media, or QC'd. Re-mastering of data from such media is done only on receipt of a request for the data.

Each standard data type is associated with a specific QC workflow. These workflows are represented diagrammatically in Figure 1. As the NHDA is the final repository of these data, data copy to LTO2 is integral to the QC procedure. Ultimately, these data will require reproduction to other online storage such as the SAN so that data is preserved without the overhead of repeated tape transcription.

The archival process can be summarised as consisting of four phases:

1. Data reception
   1.1. Allocate inventory numbers & check physical media
   1.2. Set up paper and electronic records of the inventory
2. Basic data integrity checking
   2.1. Generate summary reports and listing of data on tape
   2.2. Compare above summary to written labelling and Contractor re-mastering documentation. Identify problem media or missing data. Generate spreadsheet comparing expected with actual data.
3. Generate disc images of the data sets and write to LTO2. Check the LTO2 media and generating a report of the tar files on the tape.
4. Complete the process and transfer all QC documentation to the SAN

2 Reception of Data

2.1 INITIATING PROCEDURES FOR A NEW SURVEY

The initial step is to agree what data are to be archived. This is a process where

- BGS examines data catalogues generated by the Licensee
- Agrees the selection with the DTI
- Agrees a contract with the Licensee for the delivery of the data.

At this stage the number and names of the seismic survey(s) are known, and there is the data selection catalogue of items available.

The Seismic Data Manager is notified that a new seismic survey is to be archived. The notification will include:

- The 'company' name(s) of the survey (as used by the Licensee in the data selection catalogues and on the tape media)
- The formal DTI PON9 seismic name as defined on DEAL (this name must appear in all spreadsheets and documents generated during the archive process)
- The Name of the Licence feature that the survey is associated with
- The location on the SAN of the working folder that will hold all files generated during the archive process.
- The name(s) of the data selection catalogue(s) held in the folder.
- The estimated date of when the media will be delivered

2.2 ALLOCATION OF INVENTORY NUMBERS & CHECK OF PHYSICAL MEDIA

Every media item received into the NHDA must be given a unique identifier. As every item received will be part of a dataset, each dataset will also have a unique number assigned to it based on the year received. Thus the 'Hutton' dataset has the unique identifier N05001 as this was the first dataset to be received in 2005. Every media item belonging to this dataset is given a sequential number starting from 001. Thus the first item has the full unique identifier N05001-001, the last N05001-022.

During the inventory process the physical condition of the media should be inspected to identify damaged cartridges. Physically damaged media should not be assigned numbers and the NHDA manager informed of the problem.

The newly inventoried media should be transferred to a standard NHDA Box. Each Box can hold up to 66 IBM3590 cartridges and should be clearly labelled with dataset and item number range. A box may contain many small datasets, but splitting of datasets between different boxes should be avoided where possible. A spreadsheet itemising box numbers and their contents will be maintained and stored in W:\IM\NHDA_DEV\data.

2.3 SETTING UP PAPER AND ELECTRONIC RECORDS OF THE INVENTORY

Each item in the archive must have a corresponding NHDA data description sheet. An example is provided in Appendix 1. The most important information required is the data type and whether these are 2D or 3D datasets as these will influence which QC procedures are to be followed. The SEGD format was specifically developed for raw field data. Thus when tapes are categorised as SEGD and are then further described as 'filtered migrated' an error in labelling has probably occurred. Table 1 lists the geophysical file formats used in the NHDA and the sort of data they contain. Thus in the example above filtered migrated data is a common description of processed data volumes distributed as SEGY format files not SEGD. Procedures aiding in the identification of file format will be described later.

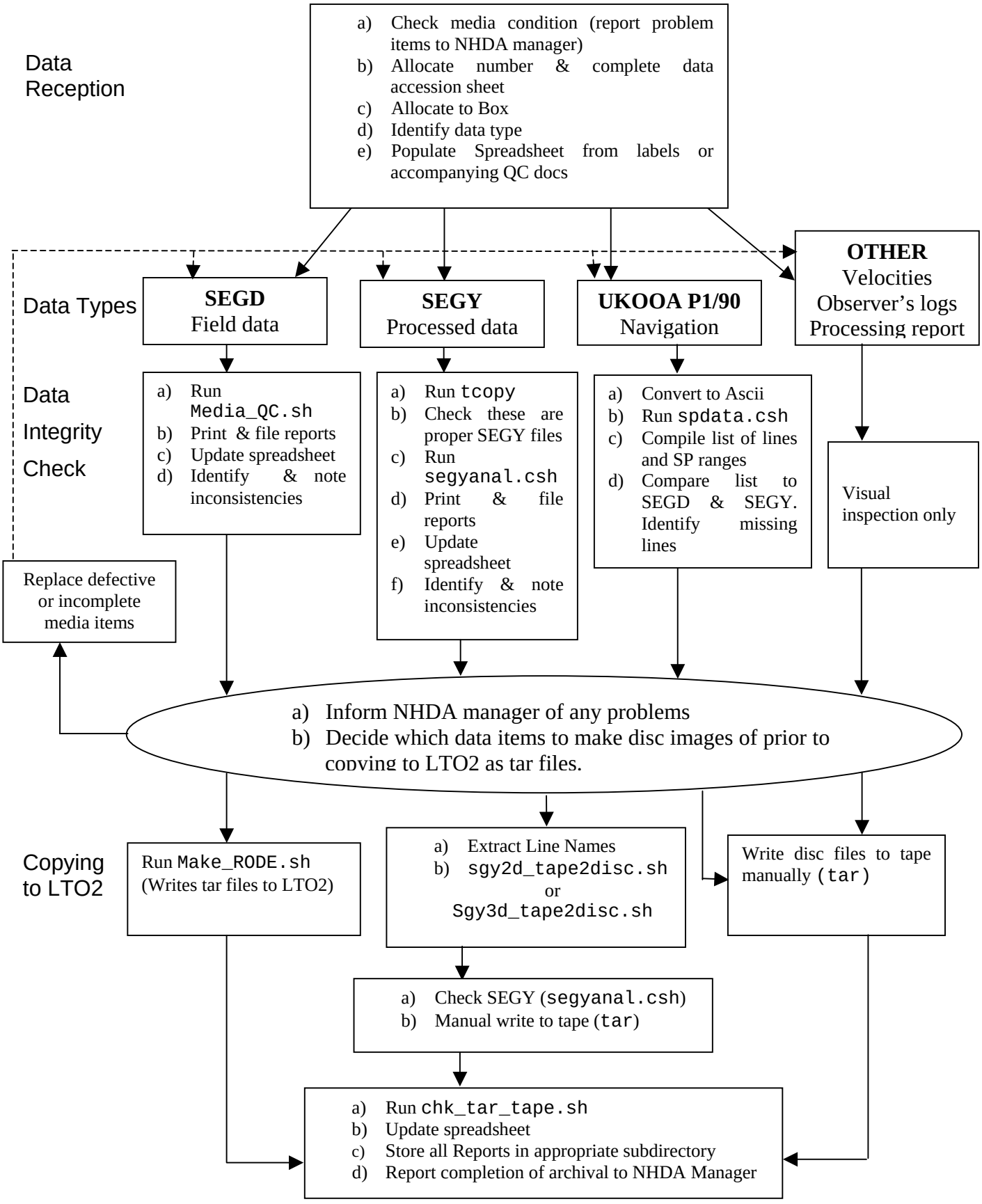
Table 1. Geophysical file formats and their contents

File Format	Possible data stored
SEGD	Field data on tape only, both 2D & 3D, several flavours exist depending on whether the data is multiplexed or demultiplexed and the numerical representation used e.g. 8015 indicates demultiplexed format using 20 bit binary numbers.
SEGD RODE	Encapsulated version of SEGD. Used to preserve blocking structure of SEGD tape files on disc.
SEGY	Filtered Migrated stack for both 2D & 3D. Also occurs in multiple flavours. Originally designed for tape but disc version is commonly used
SEGY	Raw Migrated stack, both 2D & 3D
SEGY	Raw Stack, both 2D & 3D
SEGY	On rare occasions field data for 2D, often modified with additional data such as gun signature records
P1/90	Navigation data usually in EBCDIC when written to tape but also ASCII especially in disc image
ESSOV2	A structured ASCII table used to store stacking and migration velocity data.

A physical copy of each sheet should be printed and placed in arched lever files as a reference. In addition, a photocopy of the cartridge labels should be made and placed with the data description sheet in the arched lever files. The photocopy often aids identification and may contain additional information not entered on the data description sheet.

Each archived data set should be accompanied by a QC spreadsheet generated by the geophysical contractor that transcribed the data to new media. This should list the media items and provide a description of the data on the respective media. This is important to help identify missing media items, but also to assist in instances where physical labelling of the media contains inadequate description of its contents e.g. the label of a SEGY cartridge containing over 50 files isn’t large enough to list individual line name and shot ranges.

Figure 1. NHDA processing flow from accession to completion



3 Basic check of data integrity

3.1 EQUIPMENT

A special account ID ‘nhda’ has been created on ‘mhud’, a solaris unix server which has an IBM3590 drive with ten slot autoloader and LTO2 reader and tape library attached. To achieve maximum data transfer the IBM3590 uses the IBM software driver. Additionally, Vertitas DGC’s SGYCPY has been leased for analysing and copying SEGD tapes. Landmark Graphic’s Corporation analysis tools are employed for the inspection of SEGY and UKOOA navigation data.

3.2 GENERATE SUMMARY LISTING OF DATA

Different data types require different procedures as are identified in figure 1. Each major data type is dealt with below.

3.2.1 SEGD data on tape

Specialised software called SGDCPY has been leased from Veritas DGC Ltd that analyses and copies SEGD tapes either to other tape media or to disc as RODE files. The software is in-house software designed to run on a linux PC system called Dingo that has been ported to run under Solaris. As the software is leased it has a software key that is time limited. Consequently new keys are required periodically that are issued by Veritas DGC Ltd (contact Charles_Thorns@veritasdgc.com).

3.2.1.1 MEDIA QC

Before attempting to copy the tape it is essential to check that the media is sound and that all the data that should be there are there. A shell script Media_QC.sh performs a rudimentary analysis of the SEGD tape. Appendix 2 lists the files Media_QC.sh and Media_QC.CRD and one of the output report files called N05002-6.QC. The QC file for each SEGD tape needs to be retained on the SAN as part of the NHDA documentation.

The shell script sets up report file names and handles the autoloader and invokes SGDCPY. The script is written on the assumption that the autoloader is permanently set to automatic mode, so that on unloading a tape, effected by the ‘mt’ command in the script, the autoloader looks for the next tape and loads it into the drive. The variable NHDA in the script is a string that is set to the dataset being analysed, in the example it is N05002. The script can be edited to change this. The variable I is the string containing the tape number in the ‘for’ statement. Note that these are string variables and need not increment in any order. The only constraint is that the order of the labels should correspond with the order of the physical tapes in the autoloader. Using the variables NHDA and I, report names are built that are stored within the temporary file reportname.txt along with the rptfile keyword understood by SGDCPY. Thus in the example shown, the first label in the ‘for’ statement sets I to be 02. Thus the first report name is N05002-02.QC and reportfile.txt contains one line i.e. rptfile=N05002-02.QC.

SGDCPY uses configuration or card files (.CRD extent) that informs the program which functions to perform via keywords. The file Media_QC.CRD is reproduced in Appendix 2. CRD files can include other text files, in the example the file reportname.txt generated by the script, is

included within the CRD flow. In this way the analysis for each tape is written to a separate file without the need to edit the CRD file.

SEGD files can be of various flavours; the most important difference is between multiplexed and de-multiplexed formats. Digital multi-channel seismic acquisition requires real-time measurement and recording of voltages from a large number of individual elements in a recording cable after the firing of a shot. This task is achieved by polling of each element in turn, and writing the data to tape as one large data block. The raw data records that result are called multiplexed data because the data from all recording elements are scrambled together. However, for further processing, the data samples need to be rearranged into individual channels, each channel or trace being a separate record. The latter format is known as de-multiplexed data.

This crucial difference is not automatically detected by SGDCPY. Thus it is necessary to set the keyword 'demux' in the CRD file appropriately. If this is not set properly the output will not contain a listing and error messages indicating that data is being skipped will be displayed at the terminal. An example of a QC report generated by SGDCPY on a multiplexed SEGD file is shown in Appendix 2 file listing N06001-005.QC. In multiplexed formats the data samples are not ordered by channel but by sampling order so the number of channels does not appear in the listing.

Execution time is a function of how much data there are on each tape. A cartridge with 8 Gb of data takes about half an hour. So a set of ten cartridges that's fully populated can take approximately six hours to complete.

3.2.1.2 INTERPRETING THE QC OUTPUT

The first QC issue to consider is whether the tape is physically readable. A totally corrupt tape will produce error messages. However, a large number of read errors on the QC output may indicate that the tape may become unreadable in the near future and so need to be reported. Field file numbers (FFN) are mostly shot records but often include test shots and QC files for establishing ambient noise levels, i.e. recording without actually firing. Thus FFNs will usually almost equate to or be slightly greater than the shot number range. The QC output lists the sequence of FFNs encountered and their number. The summary listing lists only the beginning and end (indicated by an * before it) of a sequence of FFNs where basic values increase monotonically. Thus the Media_QC listing in Appendix 2 shows that the sequence (SeqNum) 1 to 1018 is a series of FFNs numbered 100-117. Then 1019 to 1128 represents another set of FFNs 100-209, followed by sequence 1129 & 1130 (FFN 209 & 210 respectively). These are reported separately because of some parameter not increasing monotonically even though the FFN is increasing by one. The FFNs increase until 1244 with no FFNs missing. Overall, there are two FFN sequences present on the tape SeqNum 1-1018 (FFN 100-1117) and SeqNum 1019-2163 (FFN 100-1244). The summary listing provides sequences of FFNs, but doesn't uniquely identify the line names. Here we are reliant on either the tape label or accompanying documentation as to what lines and shot ranges to expect on the tape. The QC listing should enable us to answer the following questions:

- Is the tape readable & are there tape read errors present?
- How many FFN sequences are there on the tape?
- Are there individual missing FFNs?
- Is the range of FFNs similar to Shot ranges in the documentation?

Because of the nature of multiplexed acquisition the Media_QC report for a multiplexed tape can only report FFN ranges. De-multiplexed tapes can easily provide information as to the number of channels present. Usually, this should be constant for any one line. Although, equipment problems encountered during the survey may have required the operators to turn one or two channels off.

3.2.2 SEGY data on tape

SEGY is a standard data interchange format for processed seismic data. It too comes in many flavours including disc image. The normal SEGY file on tape can readily be identified from its blocking structure. A UNIX utility called `tcopy`, which is used for tape-to-tape copying, also lists the blocking structure. Thus a command line such as:

```
tcopy /dev/rmt/0st > N05003-001.tcopy
```

This will analyse a tape on the IBM3590 device and write the contents to the file indicated. An example `tcopy` listing is shown in Appendix 4. SEGY files always have a 3200 bytes EBCDIC record (a text block) followed by a 400 byte binary header and then followed by a sequence of records of equal length, one record per trace. Having identified that the files on tape are indeed SEGY the following script runs Landmarks GUI seismic tape analysis program:

```
csh seganal.csh
```

This program produces analyses of each file on tape, identifying trace and shot data and navigation data located in trace headers. These information as well as the EBCDIC header converted to ASCII is written to a report file with the '.log' extension. The report file should be appropriately named, i.e. with item's accession number e.g. `N05003-004.log`, saved and printed as the definitive record for the data. An example of the report is presented in Appendix 4. Both the `tcopy` output and log file should be retained and stored on the SAN as they are basic QC documents. Note that `seganal` will crash if non-SEGY files are encountered on the tape, which is why the tape must be analysed with `tcopy` first.

3D seismic can also be analysed in the same way. However, attention needs to be paid to the number of files that need to be analysed. Sometimes a 3D dataset may be written as a few SEGY files of up to 2Gb in size or a large number of files containing just one line of the 3D survey. The process outlined above still needs to be performed, but it might be undesirable to print out the full listing. Like SEGD, SEGY comes in a number of flavours. The `seganal` program can cope with many common variants where different numerical formats are employed. However, it may not handle variable trace length formats.

3.2.3 UKOOA navigation data

Navigation data is usually provided as an ASCII file in UKOOA P1/90 format. These data can be provided either on CD or tape. When written to tape they are usually in EBCDIC format and need to be converted. Run the following command to obtain the blocking structure:

```
tcopy /dev/rmt/0stb > N06001-034.tcopy
```

Listing the file reports

```
file 1: records 1 to 43: size 6400
file 1: eof after 43 records: 275200 bytes
eot
total length: 275200 bytes
```

This tape has a file with a blocking factor of 6400. To read the data, convert it to ASCII and ensure that a carriage return is present after 80 characters enter the command:

```
dd if=/dev/rmt/0stb of=N06001-034.nav ibs=6400 obs=80 conv=ascii cbs=80
```

Note the ibs (input block size) parameter is set to the same value as obtained from the tcopy listing. The other parameters in the command line are:

if= input file; of= output file; obs= output block size
conv= convert to; cbs= convert block size

Part of the output file is shown below:

```
INTERNATIONAL 1924  EUROPEAN DATUM 1950 UNIVERSAL TRANSVERSE MERCATOR
S164-89-200          101 592240.9N  62152.7W  649726 6585264  165              1
S164-89-200          105 592242.8N  62158.9W  649627 6585318  164              1
S164-89-200          110 592244.9N  622 5.9W  649513 6585379  163              1
S164-89-200          115 592247.1N  62213.1W  649398 6585442  164              1
S164-89-200          120 592249.3N  62220.6W  649277 6585505  164              1
S164-89-200          125 592251.3N  62227.4W  649166 6585564  166              1
S164-89-200          130 592253.4N  62235.1W  649043 6585625  165              1
S164-89-200          135 592255.5N  62242.4W  648925 6585683  167              1
S164-89-200          140 592257.3N  62249.4W  648812 6585736  167              1
S164-89-200          145 592259.2N  62256.7W  648694 6585790  167              1
S164-89-200          150 5923 1.3N  623 4.8W  648565 6585849  168              1
S164-89-200          155 5923 3.5N  62312.5W  648441 6585911  169              1
S164-89-200          160 5923 5.4N  62320.0W  648319 6585967  169              1
S164-89-200          165 5923 7.5N  62327.4W  648200 6586028  168              1
S164-89-200          170 5923 9.5N  62334.4W  648088 6586085  168              1
S164-89-200          175 592311.7N  62341.8W  647968 6586147  168              1
S164-89-200          180 592313.9N  62349.4W  647845 6586211  167              1
S164-89-200          185 592316.0N  62356.3W  647735 6586273  166              1
S164-89-200          190 592318.2N  624 3.4W  647620 6586336  166              1
S164-89-200          195 592320.6N  62410.8W  647500 6586404  168              1
S164-89-200          200 592322.8N  62418.2W  647380 6586468  166              1
S164-89-200          205 592324.9N  62425.3W  647266 6586530  166              1
S164-89-200          210 592327.0N  62432.9W  647143 6586591  166              1
```

The example shown represents the bare minimum of information required to locate the data. The header section usually lists the coordinate reference system (CRS) as well as other data pertinent to the survey, such as the acquisition system, ship etc. The header section is then followed by the line data. Each shot point is listed with line name, SP number, Latitude, Longitude, UTM-X,

UTM-Y & water depth. In this example, every fifth shot is listed but typically every shot is listed.

A Landmark utility '`spdata`' produces a summary listing of the navigation files simplifying the QC process. It lists the total number of lines, the first and last shot point for each line and information as the increment. The program will report sudden changes of shot value. Enter the command:

```
csh spdata.csh
```

This will run `spdata` which will present a text dialog where a report can be generated by specifying data columns for the line name and shot value as well as the number of header lines to skip. Part of the report is reproduced below:

```
   13. S164-89-210

S164-89-210          101 585535.7N  728 4.7W  588205 6533038  139              1
S164-89-210         1113 59 9 0.9N  71712.8W  597991 6558193  428              1


           101.0 to   105.0 by     4.0
           105.0 to  1110.0 by     5.0
          1110.0 to  1113.0 by     3.0

        Total # of shotpoints:     204
```

As can be seen line 13 is called S164-89-210 with a shot range of 101-1110 incrementing by 5 shots. There are no sudden or odd changes of shot value that might indicate a problem. Things that need to be checked are:

- Are all the expected lines present?
- Are the SP ranges as expected?
- Are there duplicate shot points or other problems?

3.2.4 Other data

Other data types exist such as velocity data and observers logs. Velocity data are provided as ASCII files often on CD. Presently, there is no systematic QC procedure other than visual inspection. Similarly, observer's logs, acquisition and processing reports are provided as image format and are only given a visual inspection.

3.3 COMPARE MEDIA SUMMARIES TO CONTRACTOR QC DOCUMENTATION.

Every dataset to be archived should be accompanied by a QC document generated by the contractor during the creation of the new media. It is essential to make a detailed comparison between the data extracted from the media and the QC document and the external labelling on the media. The following questions arise:

- What media items do we expect to receive?
- What lines & shot ranges do we expect to receive?
- What lines & shot ranges are on each media item?
- Are the media readable?
- Do the media contain the data expected?
- Do the SEGD and SEGY shot ranges agree (2D only)?

The contractor QC documentation and media labelling should provide the answer to the first two questions. The following two questions should be answered by the Media QC report for the SEGD data and 'seganal' reports for SEGY data. The last two questions will be answered by combining these data into an NHDA QC spreadsheet, where inconsistencies can be readily identified. An example of the NHDA QC spreadsheet is presented in Appendix 8.

To illustrate the structure of the spreadsheet it is useful to work through an example. The labelling on media item N05001-001 indicated that it should contain ten profiles with specific shot ranges. So the spreadsheet contains ten lines for this media item, one for each profile and the expected shot start and end values entered in the 'File on Label' columns. The Media_QC output listed the FFNs sequences found and these were put in the 'File found on tape' columns. Missing FFNs are indicated in the Comment column. After entering these columns and the tape format, the status flag should be set to Q, indicating that a basic QC has been completed.

Significant discrepancies, such as seen for line H-90-1031 on media item N05001-004, need to be reported to the NHDA manager who will contact the supplier for clarification or a replacement tape. Note that it is only possible to correlate line shot ranges between SEGD and SEGY for 2D profiles. This is because 3D seismic data volumes are processed into a rectangular grid of pseudo-lines and cross-lines that no longer have a simple relationship to the acquisition profiles.

4 Generating disc images of the data sets.

Once discrepancies and other media issues have been resolved, disc images of the data tapes can be created which are then written to LTO2 tapes using the UNIX tape archive (tar) utility. However, each media type has specific issues regarding the creation of disc images that is dealt with below.

4.1 SEGD

SEGD formally exists only as files on tape with a particular blocking structure. Thus SEGD can only be distributed on tape and not as disc files on CDs or portable hard drives. To permit SEGD to be stored on disc they have to be converted into RODE files. A RODE file is a disc image of the SEGD tape file that preserves the internal blocking and number format such that a new SEGD tape can be created from it that is an exact replica of the original tape.

The RODE approach has two advantages over a tape-to-tape copy. Firstly, the RODE file can be stored on disc, on tape as a tape archive (tar) file, SAN or whatever new storage medium will become available in the future. Secondly, the link between the original media archived and the copy is easily preserved, by using a suitable name e.g. `N05001-002.rode` for item N05001-002 so the process becomes self-documenting. This would be much more difficult with tape-to-tape copying as usually the new media has a much higher capacity than the old, e.g. IBM3590 holds 20 Gb of data whereas LTO2 holds 200Gb, so ten 3590s can be concatenated onto one LTO2. However, once concatenated these data will simply exist as files on tape and cannot be uniquely attributed to the original media without accompanying documentation indicating the source of the data.

Each SEGD Cartridge is encapsulated as a RODE disc file using the Veritas DGC dingo software. The RODE file is then written as a tar file to the LTO2 drive. This has a number of advantages over conventional tape-to-tape copying. Firstly, the identity of the original SEGD cartridge is preserved, so identifying specific media items is simplified. Secondly, execution times are significantly better as SUN systems are optimised for writing to and from discs. Tape-to-tape copying prevents both source and target drives from reaching their maximum read/write speeds as the target has to wait for each block of data to be read from the source, and then the source has to wait until that block has been written to the target. Where data has been written in small block sizes, transfer rates plummet as a result. Thirdly, as we ultimately intend to move the data to the SAN, disk images of the 3590 cartridges are essential.

Make RODE files

The script `Make_RODE.sh` invokes SGDCPY to create a single RODE image of each input SEGD tape. If the resultant RODE image is greater than 2Gb the program generates several extension files to ensure that no problems occur with older operating systems that cannot read files larger than that size. Because of the potential size of the files created, the script generates these one at a time, writes them to tape and then deletes them from disc before processing the next input tape. Example files are provided in Appendix 5. As with the `Make_QC.sh` script, the

operator will edit the `Make_RODE.sh` and `Make_RODE.CRD` files to properly reflect the Media Item numbers and whether the data is multiplexed or not. In general, only ten IBM 3590 cartridges should be written to one LTO2 cartridge because of their respective native capacities (20Gb and 200Gb). However, some 2D surveys may be re-mastered such that only a small fraction of the capacity of the IBM3590 cartridge is used. This may happen on some 2D surveys where data for each line is written to a separate cartridge. Obviously, we wish to minimise media storage, so it may be useful to write more than 10 to one LTO2. Unfortunately, the SGDCPY output doesn't explicitly state the size of the data found, but does provide the data needed to calculate it. The method for calculating the size is presented in Appendix 3. The script generates an output file with the extension '.rode.QC'. This is retained in case there was an error in the generation of the RODE file. It should report the same data information as the original .QC file.

4.2 SEGY

Although originally defined as a tape format, SEGY on disc is so ubiquitous that disc images of SEGY are widely accepted as a valid data exchange format in its own right.

4.2.1 2D Surveys

Files on tape have no formal name; they are simply sequential files. Whereas disc files must have names, preferably meaningful ones. Thus the first task is to create a list of line names from the EBCDIC headers generated from the SEGANAL.LOG file generated during the initial QC. Typically, multiple versions of the SEGY file are provided representing differing post-stack processing options, i.e. each line may have a raw stack, filtered stack & migrated versions. Thus when generating the line name, the type of stack section should be included.

e.g. BG106962D-001.RAW, BG106962D-001.FLT and BG106962D-001.MIG

One IBM3590 tape can contain tens even hundreds of files with their respective versions and so manually typing up a list would be very time consuming. The task can be simplified by searching for patterns in the text with UNIX commands. All EBCDIC headers start with C and the line number as can be seen in Appendix 4, N05003-004.log. The line name is contained in the line starting C02 and the last processing option in C27. Usually, the contractor uses a standard template to generate the EBCDIC header for each survey. Thus, at least within a survey, the position of the line name and processing flow should not vary between profiles. The following procedure should allow for the extraction of line names and processing from the log file. Use `textedit` to generate a file called `search.txt` containing two lines with the characters, C02 & C27 respectively. Then type the following command string:

```
fgrep –f search.txt  N05003-004.log > N05003-004.lines
```

The command filters out all lines containing the character strings within search.txt (i.e. C02 and C27) from the SEGANAL.LOG file and writes them to an output file. The output file can then be

manually edited to generate a list of file names, one per line. The following script should be edited to point to the edited lines file (listed in Appendix 6) and then executed:

```
sh sgy2d_tape2disc.sh
```

This will read the lines file and will extract each SEGY file from tape and name it accordingly. Clearly it is extremely important to ensure that the names in the line file correspond exactly to the order of files on the tape. When running this script it is a good idea to run it in a subdirectory and also to check the SEGY files on disc using `segyanal.csh`. Once it is confirmed that the files are properly named and not corrupted the following command can be executed:

```
tar -cvf  N05003-004.tar *.sgy
```

This will create a tar file of all disc image SEGY files for the volume. This can then be appended to the LTO2 volume with the associated SEGD files. If there are 53 SEGD files on an LTO2 volume then the following command will position the write head at the end of the tape:

```
mt –f /dev/rmt/2n asf 53
```

Note the use of the n switch in the device name. This will prevent the tape rewinding on execution of the command. Then the following will write the tar file to the LTO2:

```
tar –cvf /dev/rmt/2n N05003-004.tar
```

4.2.2 3D Surveys

Whereas 2D seismic profiles have unique line names in acquisition and is preserved in the final processed versions, the same is not true of 3D seismic data. The data from many profiles have been combined to form a single data volume or rectangular grid of pseudo lines and cross-lines whose relationship to the original acquisition lines is lost. Thus there is only one survey name to consider which should be taken from the EBCDIC header. However, the data volumes are often very large. Consequently, it is rare to receive only one file for the whole survey. Furthermore, it is not uncommon for 3D data volumes to be sent in the form of one file per pseudo-line. As there can be many hundred or even thousands of pseudo-lines in a 3D survey we may need to handle thousands of individual files. A script has been written that extracts each file on tape and appends a sequential number to the name. It is invoked using the following command line:

```
sh sgy3d_tape2disc.sh
```

The command should be run in a temporary subdirectory so that the files generated by this script can easily be written to the LTO2 with the tar command as shown above. The script (see appendix 7) can be edited so that a valid name for the survey, including its data type, i.e. raw, flt & mig will be created.

4.3 UKOOA

During the QC process the navigation data was already converted to ASCII and written to disc. Thus all that remains to do is to write the disc file to the LTO2 using the tar UNIX command.

4.4 OTHER

Other data is usually in ASCII or image formats provided on CD. These can be transferred to a temporary folder and archived manually using the tar UNIX command. All these files must be copied to the OPERATOR_FILES folder in the working area.

5 Completion of the archival process

Finalisation of the archival process can be summarised as follows:

- Generate listing of LTO2 and check that all files are present
- Finalise QC spreadsheet
- Transfer all QC reports to SAN area
- Inform NHDA Manager of completion of archival process

5.1 LISTING OF CONTENTS OF LTO2

Each LTO2 cartridge is bar coded and should contain a number of tar files. The following script will generate a full listing of the contents of an LTO2 cartridge:

```
sh chk_tar_tape.sh
```

The script shown in Appendix 8 generates an output file with a '.list' extension. Inherent in the process of writing to tape is the possibility of accidentally overwriting files by miss-positioning the tape head. So it is essential to carefully check that all data has been written to the LTO2. All listings should be printed out and placed in the ring binder as a permanent record.

5.2 FINALISE QC SPREADSHEET

Having checked the LTO2 listing the status and tar tape number fields in the NHDA QC comparison spreadsheet need to be entered. Those data converted to RODE files should have the Status upgraded to 'R' and those data that have simply been converted to tar files should be entered as 'T'. The QC comparison spreadsheet now has all the information needed to identify the original media number, storage box location, format, file ranges, status and Tar Tape number.

5.3 TRANSFER ALL QC DOCUMENTS TO THE SAN

All documents generated during the archival process should be retained and stored on the SAN. Each archived survey should have a corresponding directory where all Dingo QC, 'seganal' and 'spdata' reports need to be stored. Also the LTO2 contents listing should be retained in this area. These documents provide proof that the media has been checked archived properly and will aid servicing of external data requests.

5.4 ADVISE NHDA MANAGER OF ARCHIVAL COMPLETION

Once all the above tasks are complete, the NHDA manager should be informed that the survey is complete. The NHDA manager can then formally acknowledge receipt of the data in a satisfactory state so that the donor can officially apply for release of data responsibilities from the DTI.

Appendix 1. Example of inventory data sheet

National Hydrocarbons Archive Media Library

Item No:	**N04001-001**
Item Name:	**British Gas** **Area B112/11**
Date Received:	**29/04/04**
Data Source:	**Robertson**
Original Tape No:	**10 – 13**
Tape ID:	**1**

Data Description:	
Line:	**BG96-112-13-04**
Shot Points:	**10091-10922**
Files:	**91-922**
Record Length:	**3.0 sec**
Sampling Interval:	**2.0 msec**
Data Format:	**SEGD-8015**
Description:	**Field data**

Related Documentation:

Media Format:

☐ **Exabyte** ☐ **DLT** ■ **IBM 3590** ☐ **floppy**

☐ **CD-ROM** ☐ **other** _____

Comments:

Appendix 2. Example of Scripts and outputs for checking SEGD files

File Media_QC.sh

```
#!/usr/sh
#
NHDA="N05002"
for I in 02 03 04 05 06
#for I in 11 12 13 14 15 16 17 18 19 20
#for I in 21 22 23 24 25 26 27 28 29 30
#for I in 31 32 33 34 35 36 37 38 39 40
#for I in 41 42 43 44 45 46 47 48 49 50
#for I in 51 52 53 54 55 56 57 58 59 60
#for I in 61 62 63 64 65 66 67 68 69 70
#for I in 71 72 73 74 75 76 77 78 79 80
#for I in 81 82 83 84 85 86 87 88 89 90
#for I in 91 92 93 94 95 96 97 98 99 100
#for I in 101 102 103 104 105 106 107 108 109 110
#for I in 111 112 113 114 115 116 # 117 118 119 120
#for I in 121 122 123 124 125 126 127 128 129 130
do
echo rptfile=$NHDA-$I.QC > reportname.dat
SGDCPY Media_QC.CRD
mt -f /dev/rmt/0st offline
done
```

File Media_QC.CRD

```
[segdcopy]
inputdev=/dev/rmt/0st
outputdev=/dev/null
include=reportname.dat
bufsize = 8000000
eofeot = (2,1)
demux = YES
retryrmt = 10
bksz = 4000
```

File N05002-6.QC Demultiplexed sub-format 8015

```
STARTING GUARDIAN DATA SEGD_COPY P20041110    <Mon Jun 12 16:24:18 2006>

Parameters accessed from file <Media_QC.CRD>
Constants accessed from file  <>

LINE NAME = N05002-6.QC

SEGDCOPY parameters ::

tstart          0.00
tend         5000.00
delt            2.00
ntrprime           0
ntraux             0
bufsize      8000000
headertype         4
datatype           4
sampbytes         -1
mvoltlevel         0
twoscomp           0
auxchopt          -1
inittrace          1
primechan          0
eofeot             2     1
skipif             0    50
maxhdrlen       2000
spooldir
```

```
maxspool          300
demux             YES
segdform            0
segdscalar   0.000000
renumchan           NO
numchans            0
maxchans         1024
trhdrlen           -1
inendeof           NO
retryrmt           10     0     0     0
printall           NO
pseudoeof          NO
segdexhdr          -1
bytpscan            0
hdrbytes            0
limsync            -1
syncbytes    ffffff01
tmchoff             4
ntimebyt            4
tchdelta           -1
tchshift            8
limtdelt           10
limtcherr          10
tchmask      ffffff00
tshiftsw            0
ignoret1            1
blocked             0
gapped              1
dropper             0     0     0
bksz                0
chkreclim           1     1
lineid              0     0     0     0
outputdev  /dev/null
dbparfile
chkforvol         YES
chkflrec          YES
reelid              0     0     0     0
mktestdat          NO
sid                 0     0     0
projpar
image               0
rodein              0     0
inputdev   /dev/rmt/0st
memspool            0
unload              0
barcoded           NO
listfreq            0
vanguard           NO
volid              same
volid1             same
volid2             same
chsetqc           YES
mblimit             0
eotscsi             0     0
segdprint         YES
procfiles           0
tapestats          NO
conseof           YES
mblkwrite          NO

Job ran on system [mhud]

Output tape mounted on /dev/null
(driver: , drive: , serial no: )
Start input tape ?, Maximum blocksize 8000000 bytes.
Unit: /dev/rmt/0st (driver: , drive: , serial no: )

Output     FFN  Min Tr Max Tr Errors     Channels     Status          <16:24:19>
SeqNum          Length Length Hard Soft  Read  Outpt         Line = N05002-6.QC
------     ---  ------ ------ ---- ----- ----- -----  ------

     1     100     340   3220    0     0   312   264X
* 1018    1117     340   3220    0     0   312   264X
  1019     100     340   3220    0     0   312   264X
* 1128     209     340   3220    0     0   312   264X
  1129     210     340   3228    0     0   312   264X
  1130     211     340   3220    0     0   312   264X
```

```
* 2159    1240     340   3220    0     0   312   264X
  2160    1241     340   1300    0     0   312   264X  Test
* 2163    1244     340   1300    0     0   312   264X  Test


SEGD record header parameters:
Rec=000100 Fmt=8015 RL=5.1 BSI=4.00 CS=2<264p;48a>
Rec=001241 Fmt=8015 RL=2.0 BSI=4.00 CS=2<264p;48a>
Rec=001244 Fmt=8015 RL=2.0 BSI=4.00 CS=2<264p;48a>


Summary for Reel [                      ?]

 FFNbgn FFNend Inc #Recs #Her #Ser #Syn #Drop #Len
    100   1244   1  2163    0    0    0     0    3

Read latency: 1st = 218141; 2nd = 1285; min = 357; max = 4597360; tot = 545097839; num
= 677019
Write latency: 1st = 12; 2nd = 3; min = 1; max = 645; tot = 1126801; num = 573195

GUARDIAN DATA SEGD_COPY P20041110 ended NORMALLY at <Mon Jun 12 16:45:25 2006>
```

File N06001-005.QC example of multiplexed SEGD sub-format 0044

```
STARTING GUARDIAN DATA SEGD_COPY P20041110    <Wed Jun 21 16:52:56 2006>

Parameters accessed from file <Media_QC.CRD>
Constants accessed from file  <>

LINE NAME = N06001-005.QC

SEGDCOPY parameters ::

tstart           0.00
tend          5000.00
delt             2.00
ntrprime            0
ntraux              0
bufsize       8000000
headertype          4
datatype            4
sampbytes          -1
mvoltlevel          0
twoscomp            0
auxchopt           -1
inittrace           1
primechan           0
eofeot              2     1
skipif              0    50
maxhdrlen        2000
spooldir
maxspool          300
demux              NO
segdform            0
segdscalar   0.000000
renumchan          NO
numchans            0
maxchans         1024
trhdrlen           -1
inendeof           NO
retryrmt           10     0     0     0
printall           NO
pseudoeof          NO
segdexhdr          -1
bytpscan            0
hdrbytes            0
limsync            -1
syncbytes    ffffff01
tmchoff             4
ntimebyt            4
tchdelta           -1
tchshift            8
limtdelt           10
limtcherr          10
tchmask      ffffff00
tshiftsw            0
```

```
ignoret1            1
blocked             0
gapped              1
dropper             0     0     0
bksz                0
chkreclim           1     1
lineid              0     0     0     0
outputdev  /dev/null
dbparfile
chkforvol         YES
chkflrec          YES
reelid              0     0     0     0
mktestdat          NO
sid                 0     0     0
projpar
image               0
rodein              0     0
inputdev   /dev/rmt/0st
memspool            0
unload              0
barcoded           NO
listfreq            0
vanguard           NO
volid              same
volid1             same
volid2             same
chsetqc           YES
mblimit             0
eotscsi             0     0
segdprint         YES
procfiles           0
tapestats          NO
conseof           YES
mblkwrite          NO

Job ran on system [mhud]

Output tape mounted on /dev/null
(driver: , drive: , serial no: )

Start input tape ?, Maximum blocksize 8000000 bytes.
Unit: /dev/rmt/0st (driver: , drive: , serial no: )

Output     FFN  Reclen Samps Errors     Sync Time     Status          <16:52:57>
SeqNum          Bytes  Recov Hard Soft  err  ch/err         Line = N06001-005.QC
------     ---  ------ ----- ---- ----- ---- ------   ------

     1      92 2240480  4180    0     0    0      0    Tz=171
     2      93 2240480  4180    0     0    0      0    Tz=171
     3      94 2240480  4180    0     0    0      0    Tz=171 Test
     4      95 2240480  4180    0     0    0      0    Tz=171 Test
     5      96 2240480  4180    0     0    0      0    Tz=171
     6      97 2240480  4180    0     0    0      0    Tz=171 Test
*    9     100 2240480  4180    0     0    0      0    Tz=171 Test
    10     101 2240480  4180    0     0    0      0    Tz=160
    11     102 2240480  4180    0     0    0      0    Tz=160
    12     103 2240480  4180    0     0    0      0    Tz=171
    13     104 4384480  8180    0     0    0      0    Tz=171
*   23     114 4384480  8180    0     0    0      0    Tz=171
    24     115 4384480  8180    0     0    0      0    Tz=160
*  162     253 4384480  8180    0     0    0      0    Tz=160
   163     254 4367180  8148    0     0    1    >10    Tz=160
   164     255 4384480  8180    0     0    0      0    Tz=160
*  300     391 4384480  8180    0     0    0      0    Tz=160
   301     392 4384452  7214    0     0    1      0    Tz=160
*  420     511 4384480  8180    0     0    0      0    Tz=160
   421     512 4385292  8182    0     0    3    >10    Tz=160
   422     513 4384480  8180    0     0    0      0    Tz=160
*  598     689 4384480  8180    0     0    0      0    Tz=160
   599     690 4368364  5472    0     0    1      0    Tz=160
   600     691 4384480  8180    0     0    0      0    Tz=160
*  643     734 4384480  8180    0     0    0      0    Tz=160
   644     735 4333274  8084    0     0    1    >10    Tz=160
   645     736 4384480  8180    0     0    0      0    Tz=160
*  846     937 4384480  8180    0     0    0      0    Tz=160
   847     938 4386112  5189    0     0    4      0    Tz=160
```

```
   848     939 4384480  8180    0     0    0      0    Tz=160
*  957    1048 4384480  8180    0     0    0      0    Tz=160
   958    1049 4384480  8180    0     0    1      1    Tz=160
* 1176    1267 4384480  8180    0     0    0      0    Tz=160
  1177    1268 4384480  8180    0     0    0      0    Tz=161
  1178    1269 4384480  8180    0     0    0      0    Tz=160
* 1288    1379 4384480  8180    0     0    0      0    Tz=160
  1289    1380 4339416  8096    0     0    1    >10    Tz=160
  1290    1381 4384480  8180    0     0    0      0    Tz=160
* 1511    1602 4384480  8180    0     0    0      0    Tz=160
  1512    1603 4384480  8180    0     0    0      0    Tz=171
  1513    1604 2240480  4180    0     0    0      0    Tz=171
  1514     253 4384480  8180    0     0    0      0    Tz=159
  1515     254 4384480  8180    0     0    0      0    Tz=160
  1516     255 4384480  8180    0     0    0      0    Tz=159
  1517     256 4384480  8180    0     0    0      0    Tz=160
* 1556     295 4384480  8180    0     0    0      0    Tz=160
  1557     296 4138816  7722    0     0    5    >10    Tz=160
  1558     297 4384480  8180    0     0    0      0    Tz=160
* 1568     307 4384480  8180    0     0    0      0    Tz=160
  1569      95 2240480  4180    0     0    0      0    Tz=160
  1570      96 2240480  4180    0     0    0      0    Tz=158
  1571      97 2240480  4180    0     0    0      0    Tz=171
  1572      98 4384480  8180    0     0    0      0    Tz=160
* 1655     181 4384480  8180    0     0    0      0    Tz=160
  1656     182 4377992  8168    0     0    1    >10    Tz=160
  1657     183 4384480  8180    0     0    0      0    Tz=160
* 2051     577 4384480  8180    0     0    0      0    Tz=160
  2052     578 4332464  8083    0     0    1    >10    Tz=160
  2053     579 4384480  8180    0     0    0      0    Tz=160
* 2103     629 4384480  8180    0     0    0      0    Tz=160
  2104     630 4348866  8114    0     0    1    >10    Tz=160
  2105     631 4384480  8180    0     0    0      0    Tz=160
* 2153     679 4384480  8180    0     0    0      0    Tz=160
  2154     680 4338184  8094    0     0    1    >10    Tz=160
  2155     681 4384480  8180    0     0    0      0    Tz=160
* 2168     694 4384480  8180    0     0    0      0    Tz=160
  2169     695 4384480  8180    0     0    1      1    Tz=160
* 2382     908 4384480  8180    0     0    0      0    Tz=160
  2383     909 4341522  8100    0     0    1    >10    Tz=160
  2384     910 4384480  8180    0     0    0      0    Tz=160
* 2641    1167 4384480  8180    0     0    0      0    Tz=160
  2642    1168 4384480  8180    0     0    0      0    Tz=171
  2643    1997 2240480  4180    0     0    0      0    Tz=171 Test
* 2648    2002 2240480  4180    0     0    0      0    Tz=171 Test
  2649    2003 2240480  4180    0     0    0      0    Tz=158
  2650    2004 2240480  4180    0     0    0      0    Tz=159
  2651    2005 2240480  4180    0     0    0      0    Tz=171
  2652    2006 4384480  8180    0     0    0      0    Tz=160
* 2873    2227 4384480  8180    0     0    0      0    Tz=160
  2874    2228 4384480  8180    0     0    0      0    Tz=159
  2875    2229 4384480  8180    0     0    0      0    Tz=159
  2876    2230 4384480  8180    0     0    0      0    Tz=160
* 3061    2415 4384480  8180    0     0    0      0    Tz=160
  3062    2416 4384480  8180    0     0    0      0    Tz=159
  3063    2417 4384480  8180    0     0    0      0    Tz=160
* 3136    2490 4384480  8180    0     0    0      0    Tz=160
  3137    2491 4384480  8180    0     0    0      0    Tz=161
  3138    2492 4384480  8180    0     0    0      0    Tz=160
* 3217    2571 4384480  8180    0     0    0      0    Tz=160
  3218    2572 2240480  4180    0     0    0      0    Tz=171

SEGD record header parameters:
Rec=000092 Fmt=0044 RL=4.1 BSI=1.00 BPS=536 CS=6<240p;20a;1a*4>
Rec=000104 Fmt=0044 RL=8.2 BSI=1.00 BPS=536 CS=6<240p;20a;1a*4>
Rec=001604 Fmt=0044 RL=4.1 BSI=1.00 BPS=536 CS=6<240p;20a;1a*4>
Rec=000253 Fmt=0044 RL=8.2 BSI=1.00 BPS=536 CS=6<240p;20a;1a*4>
Rec=000095 Fmt=0044 RL=4.1 BSI=1.00 BPS=536 CS=6<240p;20a;1a*4>
Rec=000098 Fmt=0044 RL=8.2 BSI=1.00 BPS=536 CS=6<240p;20a;1a*4>
Rec=001997 Fmt=0044 RL=4.1 BSI=1.00 BPS=536 CS=6<240p;20a;1a*4>
Rec=002006 Fmt=0044 RL=8.2 BSI=1.00 BPS=536 CS=6<240p;20a;1a*4>
Rec=002572 Fmt=0044 RL=4.1 BSI=1.00 BPS=536 CS=6<240p;20a;1a*4>


Summary for Reel [                     ?]

 FFNbgn FFNend Inc #Recs #Her #Ser #Syn #Drop #Len
```

```
     92   2572   1  3218    0    0   15     0    0
Write latency: 1st = 13; 2nd = 5; min = 1; max = 5945; tot = 2034865; num = 852770
GUARDIAN DATA SEGD_COPY P20041110 ended NORMALLY at <Wed Jun 21 17:11:45 2006>
```

Appendix 3. Estimating the size of data content on a SEGD tape

The SGDCPY report doesn't explicitly state the size of the data on tape but does provide the information necessary to calculate it. In the example above there are 2163 records, each records holds 312 channels. Each channel is 5s in length and sampled every 2ms. Thus there are 2500 samples per channel. As each sample is recorded as a 4-byte real number, every channel requires at least 10000 bytes ignoring trace headers. The calculation is:

Let S be size each channel in bytes:

S = 4 x (trace length / sample rate) bytes

Let N be number of channels per record (312)

Let R be the number of records on the tape (2163)

Let T = Total data space

T = S x N x R

T=1000 x 312 x 2163 = 6.7Gb

A standard IBM 3590B cartridge has a capacity of 20Gb, whereas an LTO2 has a native capacity of 200Gb. At full occupancy an LTO2 cartridge can hold ten 3590s. However, with low occupancy rates as many as twenty 3590s could be contained on one LTO2.

An alternative is use the UNIX `tcopy` command to analyse the tape. This may take some time if a lot of data is present on the tape.

Appendix 4. Example of Scripts and outputs for checking SEGY files

tcopy listing showing typical SEGY tape structure

```
file 1: record 1: size 3200
file 1: record 2: size 400
file 1: records 3 to 1167: size 4240
file 1: eof after 1167 records: 4943200 bytes
file 2: record 1: size 3200
file 2: record 2: size 400
file 2: records 3 to 1071: size 4240
file 2: eof after 1071 records: 4536160 bytes
file 3: record 1: size 3200
file 3: record 2: size 400
file 3: records 3 to 1075: size 4240
file 3: eof after 1075 records: 4553120 bytes
file 4: record 1: size 3200
file 4: record 2: size 400
file 4: records 3 to 2116: size 4240
file 4: eof after 2116 records: 8966960 bytes
file 5: record 1: size 3200
file 5: record 2: size 400
file 5: records 3 to 1155: size 4240
file 5: eof after 1155 records: 4892320 bytes
file 6: record 1: size 3200
file 6: record 2: size 400
file 6: records 3 to 912: size 4240
file 6: eof after 912 records: 3862000 bytes
eot

total length: 31753760
```

Part of a 'SEGANAL' log file (N05003-004.log)

```
          L A N D M A R K   G R A P H I C S   C O R P O R A T I O N
                          TAPE ANALYSIS LOG FILE
                       Friday, June 16 06:11PM 2006



     TEXT HEADER:

C01 ELF CALEDONIA LTD - BRITISH GAS PLC    2D LINE: 106962D-001
C02 AREA : UKCS QUAD 106 - 2D LINES
C03
C04            DATASET: 9600% RAW MIGRATION
C05          ------------------------------
C06 ACQUISTION PARAMETERS:
C07 COMPANY, DATE:      GECO-PRAKLA, JUNE - JULY 1996
C08 VESSEL       :      M/V SEISQUEST
C09 SOURCE       :      1 SOURCE: FIRING EVERY 25M
C10 SP INTERVAL  :      12.5M
C11 RECORD LENGTH:      4096MS @ 2MS
C12 FORMAT       :      SEG-D 8015
C13 CABLES              1*2400M,CABLE DEPTH=6M
C14 GROUPS,GRP INT      1*192, 12.5M: OFFSET=63M (LINE 001: 150M)
C15
C16 PROCESSING PARAMETERS:
C17 ENSIGN GEOPHYSICS LTD.        JULY 1996 - MAY 1997
C18 1. SEG-D REFORMAT             2. START OF DATA DELAY (-159 MS)
C19 3. DESIGNATURE                4. 3(18)HZ LOW CUT
C20 5. 90(72)HZ(DB/OCT)ANTI ALIAS 6. RESAMPLE TO 4MS
C21 7. AMPLITUDE DECAY COMPENS.   8. FK COHERENT NOISE ATTENUATION
C22 9. ADJACENT TRACE SUMMATION   10.FK MULTIPLE ATTENUATION
C23 11.CMP SORT (96 FOLD)         12.DBS 20/240 MS
C24 13.TRACE AMPLITUDE BALANCE    14.2D DMO
C25 15.500M FINAL VEL ANALYSIS    16.2D STACK
C26 17.GUN & CABLE STATICS        18.FX DECONVOLUTION
C27 19.2D MIGRATION
C28
C29
C30 FIRST SP: 1070       FIRST CDP: 102
C31 SP RANGE: 1070       1
C32
C33 BYTE POSITIONS (FORMAT):
C34 SHOT    : 189-192 (INT)
C35 EASTING : 217-220 (INT)      NORTHING: 221-224 (INT)
C36   ''    : 225-228 (IBMFL)       ''   : 229-232 (IBMFL)
C37
C38
C39 DATUM:ED-50. UTM INTERN. ZONE 30 CENTRAL MERIDIAN 3 DEG WEST
C40 END


     BINARY HEADER:

job_id_number = 0
line_number = 2001
reel_number = 1
traces_per_record = 1
aux_traces_per_record = 0
sample_interval = 4000
original_sample_interval = 0
samples_per_trace = 1000
original_samples_per_trace = 0
format_code = 1
CDP_fold = 1
trace_sorting_code = 2
vertical_sum_code = 0
start_sweep_frequency = 0
end_sweep_frequency = 0
sweep_length = 0
sweep_type = 0
sweep_channel_trace = 0
start_sweep_taper_length = 0
end_sweep_taper_length = 0
```

```
taper_type = 0
correlated_data_traces = 0
bin_gain_recovered = 0
ampl_recovery_method = 0
measurement_system = 1
polarity = 0
vib_polarity_code = 0



     FILE INFORMATION:

tape file number = 1
no template file selected
line name = UKCSQUA
number of traces = 1165
trace header length = 240
number of samples = 1000
start time = 0 msec
end time = 3996 msec
sample rate = 4 msec
sample format = 1   (IBM Real)
trace number extracted from byte number 1, format Integer 4-Byte
  first trace number = 1
  last trace number = 1165
  trace number increment = 1
shotpoint number extracted from byte number 21, format Integer 4-Byte
  first shotpoint = 102
  last shotpoint = 1266
  shotpoint increment = 1
x coordinate extracted from byte number 73, format Integer 4-Byte
  first x coordinate =       2525.00
  last x coordinate =      15887.00
y coordinate extracted from byte number 77, format Integer 4-Byte
  first y coordinate =          0.00
  last y coordinate =          0.00
3D line number extracted from byte number 9, format Integer 4-Byte
  smallest 3D line value = 101
  largest 3D line value = 1167
3D trace number extracted from byte number 21, format Integer 4-Byte
  smallest 3D trace value = 102
  largest 3D trace value = 1266
minimum amplitude in file = -6.12531e+08
maximum amplitude in file = 7.24006e+08



********************************************************************************


 Double EOF encountered, job ended normally.
```

Appendix 5. Example of Scripts and outputs for encapsulation of SEGD files

File Make_RODE.sh

```
#!/usr/sh
#
NHDA="N05002"
for I in 2 3 4 5 6
#for I in 11 12 13 14 15 16 17 18 19 22
do
echo rptfile=$NHDA-$I.rode.QC > reportname.dat
echo pathname=/pb_d/$NHDA-$I.rode > rodemacro.dat
echo reel=$I >> rodemacro.dat
SGDCPY Hatton_RODE.CRD
mt -f /dev/rmt/0st offline
cd /pb_d
tar cvf /dev/rmt/2n $NHDA-$I.rod*
rm $NHDA-$I.rod*
cd /pa_d/nhda/dingo
done
```

File Make_RODE.CRD

```
[segdcopy]
macrofile=/pa_d/nhda/dingo/rodemacro.dat
macrokey=none
inputdev=/dev/rmt/0st
outputdev=/dev/rmt/2n
include=reportname.dat
bufsize = 8000000
eofeot = (2,1)
demux = YES
retryrmt = 10
bksz = 4000

[rode]
dest=DISC
filetype=disc
```

File N05002-6.rode.QC

```
STARTING GUARDIAN DATA SEGD_COPY P20041110    <Tue Jun 13 20:25:22 2006>

Parameters accessed from file <Hatton_RODE.CRD>
Constants accessed from file  <>

LINE NAME = N05002-6.rode.QC

SEGDCOPY parameters ::

tstart           0.00
tend          5000.00
delt             2.00
ntrprime            0
ntraux              0
bufsize       8000000
headertype          4
datatype            4
sampbytes          -1
mvoltlevel          0
twoscomp            0
auxchopt           -1
inittrace           1
primechan           0
eofeot              2     1
skipif              0    50
maxhdrlen        2000
spooldir
maxspool          300
demux             YES
segdform            0
segdscalar   0.000000
renumchan           NO
numchans            0
maxchans         1024
trhdrlen           -1
inendeof           NO
retryrmt           10     0     0     0
printall           NO
pseudoeof          NO
segdexhdr          -1
bytpscan            0
hdrbytes            0
limsync            -1
syncbytes    ffffff01
tmchoff             4
ntimebyt            4
tchdelta           -1
tchshift            8
limtdelt           10
limtcherr          10
tchmask      ffffff00
tshiftsw            0
ignoret1            1
blocked             0
gapped              1
dropper             0     0     0
bksz                0
chkreclim           1     1
lineid              0     0     0     0
outputdev  /dev/rmt/2n
macrofile  /pa_d/nhda/dingo/rodemacro.dat
macrokey   none
dbparfile
chkforvol         YES
chkflrec          YES
reelid              0     0     0     0
mktestdat          NO
sid                 0     0     0
projpar
image               0
rodein              0     0
inputdev   /dev/rmt/0st
memspool            0
unload              0
barcoded           NO
```

```
listfreq            0
vanguard           NO
volid              same
volid1             same
volid2             same
chsetqc           YES
mblimit             0
eotscsi             0     0
segdprint         YES
procfiles           0
tapestats          NO
conseof           YES
mblkwrite          NO

RODE parameters ::

dest       DISC
filetype   DISC
source     SEGY
rodetype   SCHLUMBERGER
bksz            65536
sdbhdrs    NO
leofout    YES

Job ran on system [mhud]


Start input tape ?, Maximum blocksize 8000000 bytes.
Unit: /dev/rmt/0st (driver: , drive: , serial no: )

Output     FFN  Min Tr Max Tr Errors     Channels     Status          <20:25:22>
SeqNum          Length Length Hard Soft  Read  Outpt         Line = N05002-6.rode.QC
------     ---  ------ ------ ---- ----- ----- -----  ------

     1     100     340   3220    0     0   312   264X
* 1018    1117     340   3220    0     0   312   264X
  1019     100     340   3220    0     0   312   264X
* 1128     209     340   3220    0     0   312   264X
  1129     210     340   3228    0     0   312   264X
  1130     211     340   3220    0     0   312   264X
* 2159    1240     340   3220    0     0   312   264X
  2160    1241     340   1300    0     0   312   264X  Test
* 2163    1244     340   1300    0     0   312   264X  Test

SEGD record header parameters:
Rec=000100 Fmt=8015 RL=5.1 BSI=4.00 CS=2<264p;48a>
Rec=001241 Fmt=8015 RL=2.0 BSI=4.00 CS=2<264p;48a>
Rec=001244 Fmt=8015 RL=2.0 BSI=4.00 CS=2<264p;48a>


Summary for Reel [                      ?]

 FFNbgn FFNend Inc #Recs #Her #Ser #Syn #Drop #Len
    100   1244   1  2163    0    0    0     0    3

Read latency: 1st = 219437; 2nd = 1412; min = 336; max = 4812823; tot = 552516012; num
= 677019

Input Keywords
--------------
rp66_fs_no=16838 rp66_file_no=76484898 rode_internal_form=SEGD
encap_blk_size=65536 pathname=/pb_d/N05002-6.rode

Output Keywords
---------------
firstrec=100 lastrec=1244 outcass_no=DISC-FILE outfileno=1 rp66_fs_no=16838
rp66_file_no=76484898 rode_internal_form=SEGD rl=5120 sr=4.00
encap_blk_size=65536 copytime="13/06/2006 20:25:22" num_chans=264 num_aux=48
lost_recs=0 hard_err_recs=0 soft_err_recs=0 sync_err_recs=0 read_recs=2163
file_size=1869660582 system=mhud prog_vers=SGDCPY-20041110
pathname=/pb_d/N05002-6.rode operator=JB


GUARDIAN DATA SEGD_COPY P20041110 ended NORMALLY at <Tue Jun 13 20:45:01 2006>
```

Appendix 6. Scripts used to read SEGY files to disk

sgy2d_tape2disc.sh

```
#!/usr/bin/sh
VAR=0
ITAPE="N05002-007"
for FILE in `cat $ITAPE.lines`
do
VAR=`expr $VAR + 1`
echo "File "$VAR "on "$ITAPE" will be written as "$FILE" on disc"
dd   if=/dev/rmt/0stbn of=/disk_pb_d/$FILE.sgy ibs=64k
done
```

sgy3d_tape2disc.sh

```
#!/usr/bin/sh
# script to extract multiple files from tapes
# cd /data/public/jbu
VAL=1
#
mt -f /dev/rmt/0stb rewind
#
echo Extracting FILE $VAL from tape
dd if=/dev/rmt/0stbn of=file$VAL ibs=64k
echo `ls -l file$VAL`
#
while [ ! -s file$VAL ] # test that previous file NOT zero size
do
        VAL=`expr $VAL + 1`
        echo Extracting FILE $VAL from tape
        dd if=/dev/rmt/2bn of=file$VAL.sgy ibs=64k
        echo `ls -l file$VAL`
        echo ==============================
done
echo Extraction complete
```

Appendix 7. Scripts used to check and list LTO2 tape

chk_tar_tape.sh

```
#!/usr/sh
# check output tar files
mt -f /dev/rmt/2 rewind
TTAPE="B000004"
TARGET=$TTAPE".list"
echo Report file is $TARGET
echo Report file is $TARGET > $TARGET
for I in 1 2 3 4 5 6 7 8 9 10
do
echo ================== >> $TARGET
echo TAR FILE $I ON TAPE $TTAPE >> $TARGET
echo ================== >> $TARGET
tar tvf /dev/rmt/2n >> $TARGET
done
```

Example output

```
==================
TAR FILE 1 ON TAPE
==================
-rw-rw-r--  99/10  2139095040 Jun  6 10:38 2006 N05001-01.rode
-rw-rw-r--  99/10  2139095040 Jun  6 10:42 2006 N05001-01.rode.X1
-rw-rw-r--  99/10  2139095040 Jun  6 10:46 2006 N05001-01.rode.X2
-rw-rw-r--  99/10  2139095040 Jun  6 10:50 2006 N05001-01.rode.X3
-rw-rw-r--  99/10  207836266 Jun  6 10:51 2006 N05001-01.rode.X4
==================
TAR FILE 2 ON TAPE
==================
-rw-rw-r--  99/10  2139095040 Jun  6 11:02 2006 N05001-02.rode
-rw-rw-r--  99/10  2139095040 Jun  6 11:06 2006 N05001-02.rode.X1
-rw-rw-r--  99/10  2139095040 Jun  6 11:10 2006 N05001-02.rode.X2
-rw-rw-r--  99/10  1064445310 Jun  6 11:13 2006 N05001-02.rode.X3
==================
TAR FILE 3 ON TAPE
==================
-rw-rw-r--  99/10  1790597206 Jun  6 11:23 2006 N05001-03.rode
==================
TAR FILE 4 ON TAPE
==================
-rw-rw-r--  99/10  2139095040 Jun  6 11:29 2006 N05001-04.rode
-rw-rw-r--  99/10  2139095040 Jun  6 11:33 2006 N05001-04.rode.X1
-rw-rw-r--  99/10  2139095040 Jun  6 11:37 2006 N05001-04.rode.X2
-rw-rw-r--  99/10  2056006710 Jun  6 11:42 2006 N05001-04.rode.X3
==================
TAR FILE 5 ON TAPE
==================
-rw-rw-r--  99/10  2139095040 Jun  6 11:53 2006 N05001-05.rode
-rw-rw-r--  99/10  2139095040 Jun  6 11:57 2006 N05001-05.rode.X1
-rw-rw-r--  99/10  2139095040 Jun  6 12:01 2006 N05001-05.rode.X2
-rw-rw-r--  99/10  2023877618 Jun  6 12:06 2006 N05001-05.rode.X3
==================
TAR FILE 6 ON TAPE
==================
-rw-rw-r--  99/10  2139095040 Jun  6 12:17 2006 N05001-07.rode
-rw-rw-r--  99/10  2139095040 Jun  6 12:21 2006 N05001-07.rode.X1
```

```
-rw-rw-r--  99/10  2139095040 Jun  6 12:25 2006 N05001-07.rode.X2
-rw-rw-r--  99/10  2139095040 Jun  6 12:29 2006 N05001-07.rode.X3
-rw-rw-r--  99/10  632090090 Jun  6 12:31 2006 N05001-07.rode.X4
==================
TAR FILE 7 ON TAPE
==================
-rw-rw-r--  99/10  2139095040 Jun  6 12:42 2006 N05001-08.rode
-rw-rw-r--  99/10  2139095040 Jun  6 12:46 2006 N05001-08.rode.X1
-rw-rw-r--  99/10  2139095040 Jun  6 12:50 2006 N05001-08.rode.X2
-rw-rw-r--  99/10  1690557358 Jun  6 12:54 2006 N05001-08.rode.X3
==================
TAR FILE 8 ON TAPE
==================
-rw-rw-r--  99/10  2139095040 Jun  6 13:05 2006 N05001-09.rode
-rw-rw-r--  99/10  2139095040 Jun  6 13:09 2006 N05001-09.rode.X1
-rw-rw-r--  99/10  2139095040 Jun  6 13:13 2006 N05001-09.rode.X2
-rw-rw-r--  99/10  1667278332 Jun  6 13:17 2006 N05001-09.rode.X3
==================
TAR FILE 9 ON TAPE
==================
-rw-rw-r--  99/10  2139095040 Jun  6 13:28 2006 N05001-010.rode
-rw-rw-r--  99/10  2139095040 Jun  6 13:32 2006 N05001-010.rode.X1
-rw-rw-r--  99/10  2139095040 Jun  6 13:36 2006 N05001-010.rode.X2
-rw-rw-r--  99/10  2139095040 Jun  6 13:40 2006 N05001-010.rode.X3
-rw-rw-r--  99/10  472077710 Jun  6 13:41 2006 N05001-010.rode.X4
==================
TAR FILE 10 ON TAPE
==================
```

Appendix 8. Example of NHDA QC comparison spreadsheet

N05001 ***Kerr-McGee Hutton 3D***

	Media	Store			File on Label		File found on Tape		Status	Tar Tape	File on	Q=QC made, R=Rode file created, T disc image tar
item	*(3590/CD)*	*Box*	*Line*	*format*	*Start*	*End*	*Start*	*End*	*(Q/R/T)*		Tape	*Comment* FAT=Flagged as test
001	3590	003	H-90-43	SEGD-8015	101	422	101	422	R	B0002	1	
001	3590	003	H-90-49	SEGD-8015	101	425	101	425	R	B0002	1	
001	3590	003	H-90-55	SEGD-8015	101	429	101	429	R	B0002	1	
001	3590	003	H-90-55A	SEGD-8015	101	429	101	429	R	B0002	1	
001	3590	003	H-90-61	SEGD-8015	101	433	101	433	R	B0002	1	?158 missing
001	3590	003	H-90-61A	SEGD-8015	101	435	101	435	R	B0002	1	
001	3590	003	H-90-67	SEGD-8015	101	438	101	438	R	B0002	1	
001	3590	003	H-90-67A	SEGD-8015	101	167	101	167	R	B0002	1	
001	3590	003	H-90-67B	SEGD-8015	101	438	101	438	R	B0002	1	
001	3590	003	H-90-7	SEGD-8015	101	396	101	396	R	B0002	1	
002	3590	003	H-90-73	SEGD-8015	101	404	101	404	R	B0002	2	
002	3590	003	H-90-73A	SEGD-8015	101	290	101	290	R	B0002	2	
002	3590	003	H-90-73B	SEGD-8015	101	442	101	442	R	B0002	2	
002	3590	003	H-90-73I	SEGD-8015	5101	5445	5101	5445	R	B0002	2	
002	3590	003	H-90-79	SEGD-8015	101	446	101	446	R	B0002	2	
002	3590	003	H-90-85	SEGD-8015	101	447	101	447	R	B0002	2	
002	3590	003	H-90-91	SEGD-8015	101	454	101	454	R	B0002	2	
002	3590	003	H-90-91A	SEGD-8015	101	455	101	455	R	B0002	2	
003	3590	003	H-90-97	SEGD-8015	101	353	101	353	R	B0002	3	
003	3590	003	H-90-97A	SEGD-8015	101	463	101	463	R	B0002	3	
004	3590	003	H-90-103I	SEGD-8015	5101	5465	5101	5465	R	B0002	4	
004	3590	003	H-90-114	SEGD-8015	101	468	101	414	R	B0002	4	**?415 to 468 missing**
004	3590	003	H-90-114	SEGD-8015	101	467	101	467	R	B0002	4	
004	3590	003	H-90-114I	SEGD-8015	5101	5472	5101	5474	R	B0002	4	
004	3590	003	H-90-120	SEGD-8015	101	254	101	254	R	B0002	4	**?213 to 219 missing**
004	3590	003	H-90-120A	SEGD-8015	101	395	101	395	R	B0002	4	
004	3590	003	H-90-120B	SEGD-8015	101	468	101	468	R	B0002	4	
004	3590	003	H-90-126	SEGD-8015	101	463	101	463	R	B0002	4	