

CGGVeritas (Data Services) - Transcription Reports

Table of Contents

CGGVeritas (Data Services) - Transcription Reports	1
Report Format And Layout	1
1. Report Cover	1
2. Readme Summary	2
3. Line Summary	2
4. Report Body	2
Field Definitions	2
1. Report Cover Fields	2
2. Report Body Fields	3
Non Unique Reel Ids	4
Remark Codes	5

Transcription Reports are intended to provide a concise summary of the transcription job. They are typically generated via *SEISTORE*, which is the proprietary database application used by CGGVeritas (Data Services division). It is used to store and track all related information in a seismic transcription project. Alternatively, the report may be generated from Excel.

Sources of information used to initially populate *SEISTORE* can include observer logs, existing databases, shipping transmittals, tape labels, data headers, processing reports or any combination of these and are described by the *reference documentation* field. **During the transcription process a number of fields may be discovered to be in discrepancy with the actual data. These fields will be updated to reflect the correct value (what was found on tape) and an appropriate remark entered in the database to indicate that there exists a discrepancy with the reference documentation.** An example would be the first and last record numbers on a field tape may vary from what's quoted in the observer's log. In this case the database should be corrected to contain the values found on the tape bearing in mind that the ultimate aim of the "Transcription Report" is to contain accurate information reflecting the actual data.

This document describes the general content of the report including definitions for some of the less obvious fields.

Report Format and Layout

The Transcription Reports will be provided as a PDF file (unless specified otherwise) and are generated by CGGVeritas's *SEISTORE* database or via Excel.

1. Report Cover

The Report Cover generally contains information that is constant (static) for the job. It is divided into 3 main sections, Job Details, Survey Details and Media Details, with the appropriate field values displayed within each section. Static fields are categorised into 1 of these 3 sections according to the following definitions;

Job Details: Contains fields related to the job such as the client, job number, job description and transcription statistics.

Survey Details: Contains fields related to the survey. For field data this will reflect the survey name, year, type, acquisition company and some basic acquisition details. For processed data this will reflect the processing project name, year, processing company and basic processing details.

Media Details: Contains a summary of the input and output media including type, length or capacity, brand and data format.

Definitions for some of the less obvious fields that appear in the Cover Worksheet can be found in the section "Field Definitions".

Note that some of the fields that appear in the Report Cover, may contain more than one value such as Survey Name, Input Media Type, Input Media Capacity etc. For these fields, the message "Multiple {field name} – Refer to Database Export File For Further Details" will be displayed. The user can then view the Database Export File (Excel Spreadsheet) if they want to determine the different values for this field.

2. *Readme Summary*

Most jobs will include a text summary of the requirements of the job – this file will also appear as *readme.txt* in the Reports directory.

3. *Line Summary*

Most jobs involving concatenated field data will include a line summary showing the location of each line along with the file and SP ranges (if known).

4. *Report Body*

The Report Body generally contains those fields that are not constant (non-static) for the job. A separate column will appear for each field to be listed in the Report Body. These fields that appear in the Report Body may vary depending upon the data transcribed. For example for post stack data the field “data kind” is usually included to help differentiate different processing types (Raw Stack, Filtered Migration etc...).

Each **Row Item** in the Report Body represents a single *SEISTORE* database record, where CGGVeritas’s definitions for the fields Reel ID, Input File Number and Input Logical File Number determine when a database record is created.

REPORT BODY – ROW ITEMS

A separate Row Item (database record) will appear in the Report Body for each unique combination of the fields *reel id*, *input file number* and *input logical file number*. CGGVeritas’s definitions for these fields are;

The *reel id* is the original input reel number. In instances where the reel number is not unique CGGVeritas will apply a suffix in order to yield a unique reel id (see section “non-unique reel ids” for further details).

The *input file number* will increment for each physical file (delimited by an EOF mark) on the input reel. Note that the *input file number* field is only incremented when an EOF mark has been used as a data set delimiter and that data sets that contain multiple EOF marks as part of the seismic format are not considered separate files (e.g. SEG-D demultiplexed formats).

The *input logical file number* will increment for each unique combination of *line number* and *kind of data* contained within an *input file number* (physical file).

The Report Body’s Row Items (database records) are usually ordered by Output Media ID, Output File Number and Output Logical File Number.

Definitions for most of the fields that appear in the Report Body can be found in the section “Field Definitions”.

Field Definitions

1. *Report Cover Fields*

No. Of Copies

Indicates how many copies of the data were requested. Note that different reports will be generated if different output media have been requested (i.e. one copy to 8mm and another to 3590), therefore the number of copies represents the number of copies made to the same media type.

No. Input Reels Transcribed

This number represents the total number of input tapes transcribed to the final output. Note that all input reels are entered into *SEISTORE* as per the “Reference Documentation” and that some reels may not be transcribed (see below for further details).

No. Reels Not Transcribed

This number indicates the number of input reels not transcribed. Input reels may not be transcribed due to the reel being blank, missing or labelled as “do not process”.

No. Of Output Media

This indicates the total number of output tapes generated. Note if multiple copies have been requested to the same output media type, that this number will reflect the total number of outputs including the duplicate sets.

No. Output Boxes

If the output tapes have been shipped in boxes, CGGV-DS will assign each box a unique box number. This number reflects the total number of boxes shipped.

Reference Documentation

Refers to the main source of information used to populate the *SEISTORE* database fields. Note that in some instances more than one reference source may be required.

2. Report Body Fields

Note – *SEISTORE* provides for different report layouts, hence not all fields may be present.

Reel Id

The reel id is the original input reel number. In instances where the reel number is not unique CGGV-DS will apply a suffix in order to yield a unique reel id (see section “non-unique reel ids” for further details).

Do not process, missing and blank reels will appear in the Report Body but are distinguished from other tapes by the appropriate code in the *remark* field (see section “Remark Codes” for further details). These reels can also be identified by the fact that the *Output Media Id* field will be empty.

Barcode

If the input reels have barcodes associated with them, this information will be databased and presented in the report.

Input File Number

The *input file number* will increment for each physical file (delimited by an EOF mark) on the input reel. Note that the *input file number* field is only incremented when an EOF mark has been used as a data set delimiter and that data sets that contain multiple EOF marks as part of the seismic format are not considered separate files (e.g. SEG D demultiplexed formats).

Input Logical File Number

The *input logical file number* will increment for each unique combination of *line number* and *kind of data* contained within a *input file number* (physical file). For example a single physical SEG Y file may contain the stack and migrated data for an individual line, hence this would appear as 2 Row Items in the Report Body and the Input Logical File Number would be 1 and 2.

Input Box Number

If the original tapes have been shipped in boxes with box numbers assigned, these will be entered in the database. This is an optional field.

Kind Of Data

This *kind of data* (data type) field is generally used to distinguish different processing sequences applied to the data.

Output Media Id

The *output media id* is the unique id assigned to the output tapes generated during the transcription process. Unless instructed otherwise CGGV-DS will apply the following *output media id* convention;

AU3{job number}-XXX{A}

Where;

- Job number is a unique 4 digit number assigned by CGGV-DS to each transcription job.
- XXX is a 3 digit number which starts at 001 for each job number.
- A is a suffix letter appended if more than one copy to the same output media type has been requested. The first set will be assigned the letter A, the second B etc...

The above conventions ensure that all outputs generated by CGGV-DS (world wide) are unique.

Output File Number

The *output file number* will increment for each physical file (delimited by an EOF mark) on the output. Note that the *output file number* field is only incremented when an EOF mark has been used as a data set delimiter and that data sets that contain multiple EOF marks as part of the seismic format are not considered separate files (e.g. SEG D demultiplexed formats).

Output Logical File Number

The *output logical file number* will increment for each unique combination of *line number* and *kind of data* contained within a *output file number* (physical file). For example a single physical SEG Y file may contain the stack and migrated data for an individual line, hence the Output Logical File Number would be 1 and 2.

Output Box Number

If the output media is being boxed, CGGV-DS will assign each box a unique number. The convention used will ensure that all boxes shipped by CGGV-DS (World Wide) will have been assigned a unique box number.

Records / CDPs Dropped

This field will refer to the number of records / CDPs not recovered (dropped) during the transcription process. This may occur if the original tape is in poor physical condition. For field and pre-stack data sets this will generally represent the number of complete shot records dropped while for post stack data it will reflect the number of unrecoverable CDPs.

Hard Error Count

Is the total number of records read with hard errors. Hard errors are read errors reported back from the tape drive/formatter and can be due to a wide range of possibilities. It implies that the data has been corrupted however in some instances the geophysical integrity of the data may have been maintained (i.e. a hard error does not always imply the data is bad). This can only be determined by visually inspecting a display of the affected data.

For post stack data this will generally represent the number of CDPs containing hard errors, while for field and pre-stack data this will represent the number of records that contained a hard error. If the value of this field is 0 then the original data has been read error free, while values greater than 0, when compared with the number of records read, yields a useful qualitative measure of the tapes condition.

Soft Errors Count

Is the total number of records affected by soft errors and/or sync errors. For post stack data this will generally represent the number of CDPs containing soft errors, while for field and pre-stack data this will represent the number of records that contained a soft error. A soft error is defined as a data sample, which did not conform to the appropriate numerical format. This check is only applicable to certain formats such as SEG Y, DISCO and SEG D (2048 and 0048).

Records/CDP Read Count

The total number of records processed. For post stack data this will generally represent the number of CDPs read, while for field and pre-stack data this will represent the number of records read.

File Size

This field should reflect the size of the file transcribed in bytes.

Remarks

This field is used to describe any error, anomalies or discrepancies encountered during the transcription process. CGGV-DS have developed a set of Remark Codes to standardise the way we report errors or discrepancies. For further details see section "Remark Codes". This field is only displayed when not empty.

Non Unique Reel Ids

In the advent of non-unique input reel Ids occurring within a job, CGGV-DS will apply a suffix to the reel id to ensure it's uniqueness. The convention used by CGGV-DS is for a suffix consisting of an underscore followed by a numeric character. The numeric will commence at 1 and increment for each subsequent reel found with the same initial reel id. For example if the reel id "1032" occurs three times within a single transcription project then their reel ids will become "1032", "1032_1" and "1032_2".

For input reels to which a suffix has been applied, CGGV-DS will either apply a new “reel id” label or manually record the new id on the existing label.

The adoption of this convention overcomes many limitations associated with having non-unique reel ids but still enables the client to easily identify the original reel id. This may be important when relating the data back to the original support data.

Remark Codes

CGGV-DS have developed a set of Remark Codes to standardise the way we report errors or discrepancies encountered during the transcription process. A list and definitions for the Remark Codes can be found on the 2nd page of the transcription report (immediately following the cover page).

The use of codes was adopted for the following reasons;

- To standardise how various types of errors or discrepancies were being reported.
- The use of standard codes enables accurate queries and statistics to be calculated in the future.
- To reduce wordy explanations.

The Remarks field in the transcription report may include several comments or error codes and will only be displayed when not empty. Each code and it's associated comment is delimited by square brackets (e.g [SY][DOC – reel labels used]).