

Guardian Data - Transcription Listings

Table Of Contents

Guardian Data - Transcription Listings.....	1
Naming Convention For Transcription Listings	1
Transcription Listing Format.....	2
1. Multiplexed Input.....	2
2. Demultiplexed Input.....	4
3. Reel Summary.....	5

Guardian Data's transcription software generates a "transcription listing", which depicts the data that has been transcribed and any error encountered. The "transcription listings make use of a proprietary feature called "print reduction", which significantly reduces the amount of print, making the task of identifying potential problems or errors considerably easier.

Print Reduction

Print reduction is identified by an asterisk in the first column, signifying that one or more preceding lines have been removed as the content may be derived intuitively.

Print lines are removed only if:

- the record increment follows the currently defined pattern (initially expects an increment of 1, but will 'learn' if the increment is constant for 5 successive entries).
- There are no errors as identified by the hard, soft, sync, timing channel error or missing counters.
- There is no change in any of the other values (e.g. record length, samples recovered, channels read, etc.).

In addition, a one-line summary is created for each input reel. Guardian Data's transcription programs will automatically invoke print reduction on the listing file unless specifically turned off.

Naming Convention For Transcription Listings

The name adopted for the transcription listings will vary depending upon the data type (post stack versus field), the condition of the tapes and whether the listing represents the original transcription process or is a QC of the final output. In general the following conventions are used;

Transcription Of Original Tapes

- For field data on good quality tapes a separate listing will be generated for each line, with the line number followed by the .lst extension, used as its filename.
- For field data on poor quality tapes (sticky) a separate listing will be generated for each input reel, with input reel id followed by the extension .lst, used as the filename.
- For post stack data a separate listing will be generated for each input reel, with input reel id followed by the extension .lst, used as the filename..

QC Of Final Outputs

- A separate listing is generated for each output reel, with the output media id followed by the extension .qc, used as the filename. These listings are optional and may not be generated for every project.

Transcription Listing Format

The section of the list file, which details the data transcribed, will be set out in one of two formats depending on whether the input data is multiplexed or demultiplexed. A summary section then follows this. Below is a description and example of the two different listing file formats.

1. Multiplexed Input

The following “transcription listing” is an example of output from a job which processed a multiplexed data set (in this case, it was SEGB).

```

Start input tape D3584F-0004N

Output      FFN  Reclen Samps Errors      Sync Time      Status      <06:56:32>
SeqNum      Bytes  Recov Hard Soft   err  ch/err      Line = /PROBLEM/test
-----
      1  7016  803840  2560      0      0      0      0
*  40  7055  803840  2560      0      0      0      0
      7056      0      0      0      0      0      0      Short <Dropped>
  41  7057  803840  2560      0      0      0      0
      7058      0      0      0      0      0      0      Short <Dropped>
      7059      0      0      0      0      0      0      Short <Dropped>
  42  7060  803840  2560      0      0      0      0
  43  7061  803840  2560      0      0      0      0
      7062      0      0      0      0      0      0      Short <Dropped>
  44  7063  803840  2560      0      0      0      0
  45  7064  803840  2560      9      7      0      0

                Reel FFNbgn FFNend Inc #Recs #Her #Ser #Drop #Len
D3584F-1992N   7016   7064   1    45    0    0    4    0

```

The columns are as follows:

Output SeqNum

Identifies the logical sequence of input “records”, starting from 1. If preceded by an asterisk (*), then this indicates print reduction has been invoked, and that all intermediate sequence numbers also exist without anomalies in remaining fields.

FFN

Identifies the record ID as per the byte location selected by the operator. For field data the field record number is usually selected while for pre-stack the gather number and for post stack the cdp number.

Reclen Bytes

For each input record, this represents the length of the data block in bytes.

Samps Recov

For each input record, this represents the number of samples recovered for each trace. Note that, in the usual case where timing information is used to locate each sample #, this actually represents the implied number of samples, taking into account the position of the first and last sample.

Errors Hard

Represents the number of hard errors encountered for each input record - i.e. the number of tape blocks which encountered a read error. 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 and this can only be determined by visually inspecting the data).

In the general case of gapped data (one header block and one data block – SEG B), a value of 1 would indicate an error in either one or the other, while a value of 2 would indicate an error in both

Strictly Confidential
Errors Soft

Represents the number of soft errors encountered for each input record - i.e. the total number of samples, which did not conform to the appropriate numerical format. Note that this normally only applies to formats which use floating point numbers, where the value of the sample exceeds the permitted range.

Sync err

Guardian's software incorporates an advanced data recovery algorithm (termed "re-synchronisation") when copying bad data. This will frequently enable the program to recover valid data beyond corrupted data samples. Sync Errors indicate that this recovery technique has been invoked.

Time ch/err

Represents the number of timing channel errors (for those formats, which contain timing information). The presence of many timing channel errors (e.g. ">10") will impact the re-synchronisation algorithm's, chances of successfully recovering the data. See Sync Err above for further details.

Status

A free format textual field providing comments such as:

Input > Output	(input samples greater than output trace length).
<Dropped>	(record is dropped due specified reason - e.g. short)
No HDR	(no header found).
Bad=xxxx	(identifies first resync event).

2. Demultiplexed Input

The following list is an example of output from a job, that processed a demultiplexed data set (in this case, it was SEG Y).

Start input tape SW78Y-DMX50

Output SeqNum	FFN	Min Tr Samps	Max Tr Samps	Errors Hard Soft	Channels Read Miss	Status	<14:02:26> Line = /PROBLEM/test
1	3564	2048	2048	1 0	96 0		
2	3565	2048	2048	0 0	96 0		
3	3566	2050	2050	2 0	1 95		
4	3534	1908	1908	0 0	1 95		
5	3566	2048	2048	1 0	94 2		
6	3567	2048	2048	0 0	96 0		
* 45	3606	2048	2048	0 0	96 0		
46	3607	1899	2048	1 0	96 0		
47	3608	2048	2048	0 0	96 0		
* 51	3612	2048	2048	0 0	96 0		
52	3613	1913	2048	1 0	96 0		
53	3614	2048	2048	0 0	96 0		
54	3615	2048	2048	0 0	96 0		
55	3616	2048	2148	1 0	96 0	Input > Output	
56	3617	2048	2048	0 0	96 0		
* 68	3629	2048	2048	0 0	96 0		
69	3630	2048	2053	1 0	96 0	Input > Output	
...							
...							
136	3693	2048	2360	1 0	96 0		
137	3694	2048	2048	2 0	96 0		
* 146	3703	2048	2048	0 0	96 0		
147	3704	2048	2048	1 0	96 0		
* 151	3708	2048	2048	0 0	96 0		
152	3709	120	2048	3 0	96 0		
153	3710	1153	2235	12 5	48 48		
Reel FFNbgn FFNend Inc #Recs #Her #Ser #Drop #Len							
SW78Y-DMX50 3564 3710 1 153 90 1 0 41							

The columns are as follows:

Output SeqNum

Identifies the logical sequence of input “records”, starting from 1. If preceded by an asterisk (*), then this indicates print reduction has been invoked, and that all intermediate sequence numbers also exist without anomalies in remaining fields.

FFN

Identifies the record ID as per the byte location selected by the operator. For field data the field record number is usually selected while for pre-stack the gather number and for post stack the cdp number.

Min Tr Samps

For each input record (shot record), this represents the shortest trace length - measured in number of samples.

Max Tr Samps

For each input record (shot record), this represents the longest trace length - measured in number of samples.

Errors Hard

Represents the number of hard errors encountered for each input record - i.e. the number of tape blocks which encountered a read error. In the general case of one trace per tape block, this will represent the number of traces, which were read with 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 and this can only be determined by visually inspecting the data).

Errors Soft

Represents the number of soft errors encountered for each input record - i.e. the total number of samples, which did not conform to the appropriate numerical format. Note that this normally only applies to formats which use floating point numbers, where the value of the sample exceeds the permitted range. Note also that the listing does not identify the number of traces, which contain the errors - i.e. they could all be in one trace or spread out through several.

Channels Read

Represents the number of channels (traces) read for this input record.

Channels Miss

Represents the number of channels (traces) expected but not read for this input record. Note that a trailing "X" indicates that there are more channels available in the record than were selected by the prime and auxiliary trace selection parameters (for SEG D). If this count is positive, it implies that there are channels missing from those selected. If the count is negative, it implies there are additional channels present from those selected.

Status

A free format textual field providing comments such as:

Input > Output (input samples greater than output trace length).

Tz=xxxx (value of start time, if present – e.g. SEG D)

3. Reel Summary

At the completion of each reel, a one line summary is output, as per the following example:

Reel	FFNbgn	FFNend	Inc	#Recs	#Her	#Ser	#Drop	#Len
SW78Y-DMX50	3564	3710	1	153	90	1	0	41

These values have the following meanings.

FFNbgn

The first record for this input reel.

FFNend

The last record for this input reel.

Inc

The increment value between records.

#Recs

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.

#Her

The total number of records affected by hard errors (not the same as the total number of hard errors). 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.

#Ser

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).

#Drop

The total number of records which were dropped.

#Len

The total number of times the Length fields (Min Tr/Max Tr) varies from the previous line. Note that this is not the same as the number of records with length errors (for many formats, the program cannot determine what the proper length is).