9th Biennial International Conference
& Exposition on Petroleum Geophysics

"HYDERABAD 2012"

P-174

Establishing Standards for Transcription of Seismic Data: Process, QC and Learning

Rashmi Bhangale*, Mandadapu Kishore, Lakshmi Ramana, *RIL*

Summary

Vast quantities of seismic data are collected over the life cycle of an E&P company from scouting, through Exploration, Development and Production. Conventionally this data is stored and maintained on 3590 tapes which grow in to large numbers (>100,000) in line with size and age of company. With 8-10 year shelf life, these large tape- libraries need expensive facilities and operational maintenance. Need for Transcription is due to factors like Growth and Space Requirement, Media Deterioration, Managing End of Life Technologies, Business Continuity and Disaster Recovery. This work has developed a Standard Operating Procedure for Transcription QC, which is not available in the industry. The paper presents the SOP developed and tested for a large and comprehensive transcription project of an entire seismic repository. This methodology is proposed to be considered for adoption as a reference framework for standardization of seismic Transcription QC at E&P companies.

Keywords*: Seismic Tape Transcription, Standard Operating Procedure (SOP), Data Formats, Bit to Bit Copy, MD5 check, SEG-D, SEG-Y, Navigation Maps, Time slices, QC*

Introduction

Huge volumes of SEG-D and SEGY data is generated from Acquisition, Processing and Interpreatation. New generation of processing sequences involving pre-stack gathers create multiple volumes of data. With increasing focus on 3D, 4D and special processing, the volumes of data tapes increase exponentially. Considering their business value, a complete Disaster Recovery set is maintained by all E&P companies. The regulatory requirements compound the volumes. This project addresses a substantial volumes of data tapes ~100,000 covering the entire seismic data repository to be Tranascribed and QCed.

This is a pioneering initiative of this scale and comprehensiveness for 3590 media in the E&P industry. This large volume data transcription necessitated meticulous planning and developing of the following procedure:

i. Objectives and Requirements
ii. SMACed Scope of Work
iii. Selection of Output Media
iv. Project execution, infrastructure and facilities
v. SOPs for Transcription QC

The various considerations and adopted solutions are presented.

Objectives and Requirements

The project was conceived to be accomplished in 18 months within firm budgets. Creating a consistent and complete set of new data tapes – 2 sets for operational and DR purpose is scoped. Correctness – described by bit-to-bit copy and Completeness – described as coverage of every file in the repository are identified as requirements. Correctness and Completeness are designed to be demonstrated using measured metrics through the work-flows.

SMACed Scope of Work

Business requirements during the Project Phase and subsequent Operations Phase of the seismic tape library are covered.

i. Cleaning and reconditioning of the input tapes/

Reliance Industries Ltd. Petroleum Business (E&P), Reliance Corporate Park, Building 10BFF, Thane Belapur Road, Ghansoli, Navi Mumbai – 400701, Maharashtra, India.
Email – rashmi.bhangale@ril.com

cartridges on need basis

ii. Bit-to-Bit Copying in 2 sets

iii. QC and Data validation

iv. Meta data updation in a database

v. Building an Online Intranet Portal (.NET) with backend Database (SQL Server) for monitoring the project progress and subsequent Operational use.

vi. Specific (S), Measurable (M), Achievable (A) and Complete (C) metrics are set for the above.

Selection of Output Media

Data needs to be transcribed from current 3590 media to a suitable output media. Criteria considered are: i) Robust technology, ii) Shelf life, iii) Transfer speeds, iv) Industry compatibility, and v) Write Protection. IBM WORM JB/JX 700 GB 3592 media is chosen considering the excellent data transfer (read/write) rates i.e. 2.5 GB per minute, and ~10 year shelf-life. It was confirmed to be compatible with the systems with acquisition and processing vendors. The WORM media (Write Once Read Many) will ensure data security and avoid accidental overwrite by users.

Project Execution, Infrastructure and Facilities

Tape issue and return for the Transcription project will involve continuous data tape handling. To address this Tape Library and Transcription Area are collocated in the same premises. 24x7 work flows of the project required – separate server room, power with redundant UPS, air-conditioning, security, staff-workplace and other services. Capacity planning is meticulously done to support the resource sizing for the entire project. The facilities are designed and developed before project mobilization. The facility now supports 30 tape drives, servers, storage, QC stations and data-handling area to cycle ~8000 input data tapes per month.

SOP for Transcription QC

Pioneering work has been accomplished in developing a very comprehensive and robust SOP for achieving the 'Correctness' and 'Completeness' criteria. The SOP has been implemented and is being operated with 100% reliability through systematic Project Management.

i. Data Scheduling

ii. Input Data Preparation

iii. Transcription

iv. Data validation & verification (Input, Output)

v. Output Creation

vi. Acceptance QC

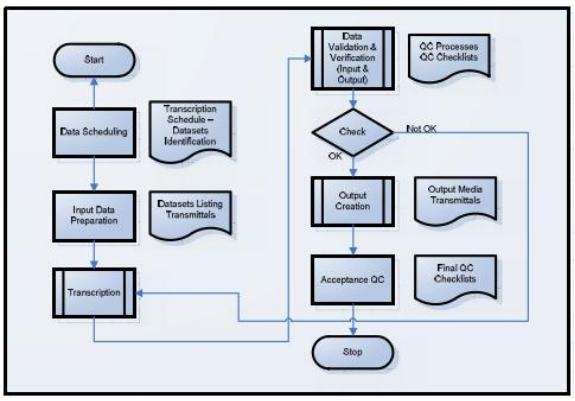


Figure 1 Transcription & QC Workflow

QC Requirement	Method Adopted
Ensuring Bit to Bit Copy	MD5 check
Trace Data Validation	NT & QC
Operational Data Loading & Analysis	EBCIDIC Header
Coverage of All Lines in a Suvery	Navigation Maps
Ensuring complete coverage of data	Timeslices (Top, Middle, Bottom)
Data Readability from Transcribed Output tapes	Read output media through Applications like Ladmark Promax & CGG Geocluster

Table 1 Quality requirement and method adopted in Transcription QC of Seismic data.

Three levels of validation and verification followed by a final acceptance QC is the pivot of the new SOP developed in this project. Salient aspects of this are elaborated.

MD5 Checksum ensuring Bit-to-Bit copy

The MD5 Message-Digest Algorithm is a widely used cryptographic hash function that produces a 128-bit (16-byte) hash value, something like a fingerprint of the file. There is a very small possibility of getting two identical checksums of two different files. This feature can be useful both for comparing the files and their integrity control. An MD5 hash is typically expressed as a 32-digit hexadecimal number.

MD5 checksums are produced during copy from input 3590 tape to disk. The input is then concatenated together to be output to the 3592 output media. MD5 checksums are also produced while copy from disk to output 3592 media.

These two Input and Output list files are comapred for their MD5 checksums. If they match then both the input and output tapes are a perfect copy.

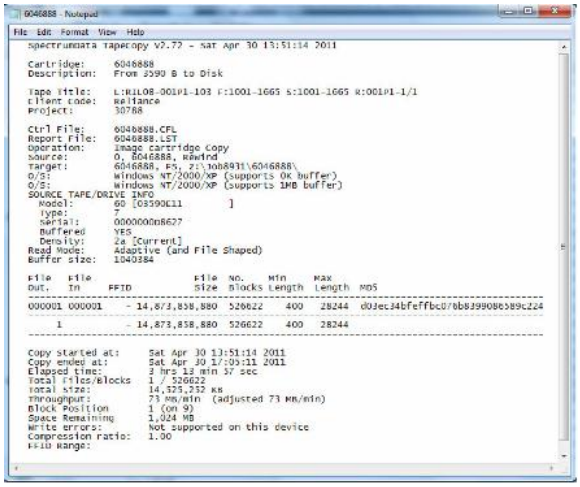


Figure 2 MD5 Check Input List file

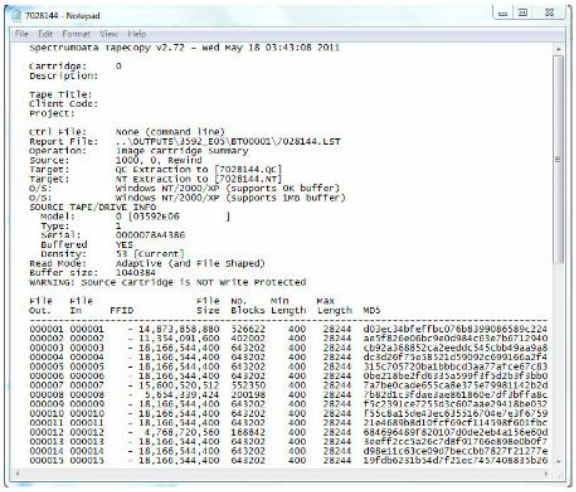


Figure 3 MD5 Check Output List file

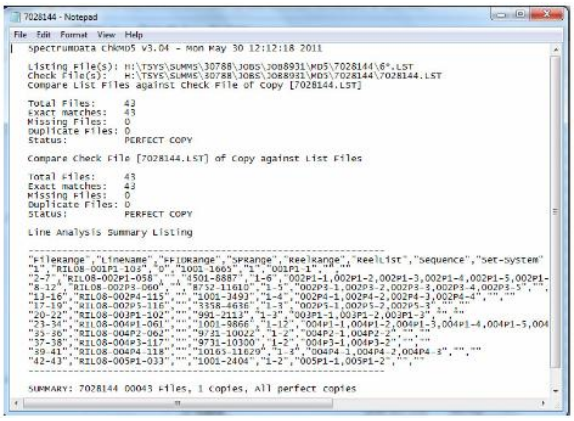


Figure 4 MD5 Check Compare file

Trace Data Validation for Gathers – NT & QC

NT file for SEG-D and SEG-Y Gather data contain display of near trace data for every FFID or Shotpoint. Through the SEGY view application these NT files are displayed and it is ensured that every FFID value gets displayed and there are no gaps or spikes in the data.

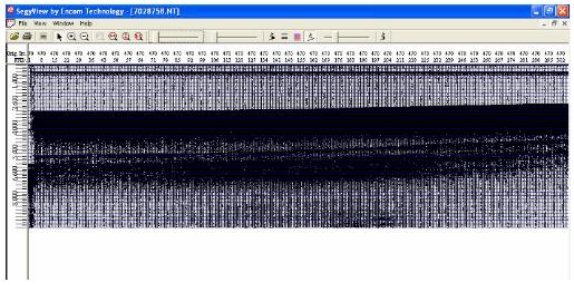


Figure 5 NT file

QC file for SEG-D and SEG-Y Gather data contain all trace data for every 1000th file of a particular line. For example. If there are 4000 files for a particular line then there will be 5 FFID's/Shotpoints with their respective trace displays for 1st, 1000th, 2000th, 3000th & 4000th file.

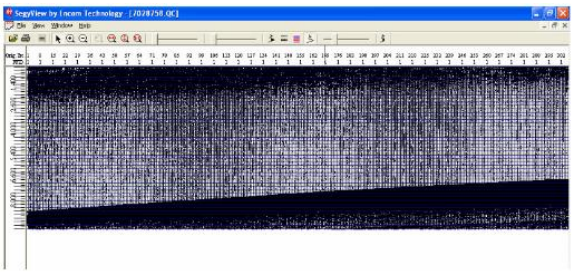


Figure 6 QC file

Timeslices (Top, Middle & Bottom)

Timeslices generated at the Top, Middle and Bottom of the data ensure complete coverage and no data gaps.

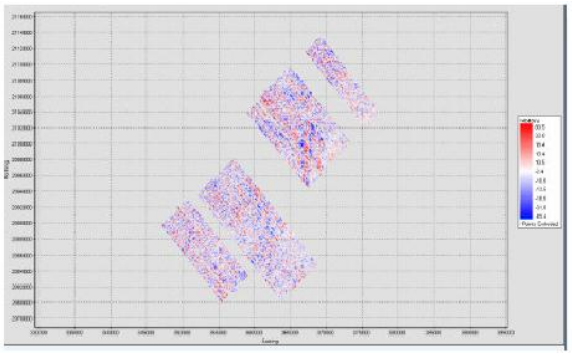


Figure 7 Timeslice for Top

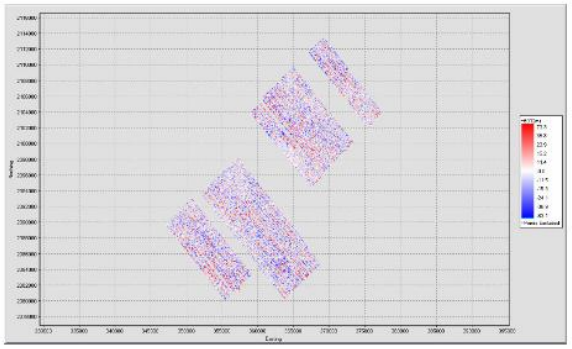


Figure 8 Timeslice for Middle

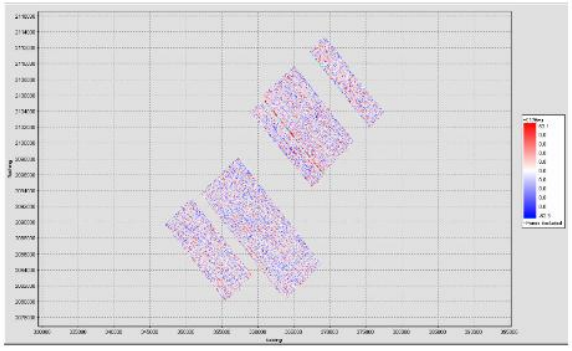
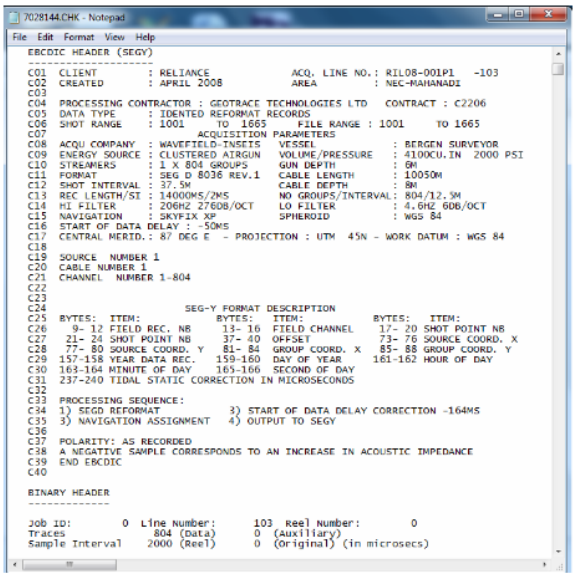


Figure 9 Timeslice for Bottom

EBCIDIC Header for SEGY data

The EBCIDIC Headers in pdf format for all SEGY data will help for Operational data loading and data analysis in the future.



7028144.CHK - Notepad

File Edit Format View Help

```
EBCDIC HEADER (SEGY)
--------------------
C01  CLIENT          : RELIANCE              ACQ. LINE NO.: RIL08-001P1   -103
C02  CREATED         : APRIL 2008            AREA         : NEC-MAHANADI
C03
C04  PROCESSING CONTRACTOR : GEOTRACE TECHNOLOGIES LTD   CONTRACT : C2206
C05  DATA TYPE       : IDENTED REFORMAT RECORDS
C06  SHOT RANGE      : 1001     TO  1665      FILE RANGE : 1001     TO 1665
C07                          ACQUISITION PARAMETERS
C08  ACQU COMPANY    : WAVEFIELD-INSEIS   VESSEL              : BERGEN SURVEYOR
C09  ENERGY SOURCE   : CLUSTERED AIRGUN   VOLUME/PRESSURE     : 4100CU.IN  2000 PSI
C10  STREAMERS       : 1 X 804 GROUPS     GUN DEPTH           : 6M
C11  FORMAT          : SEG D 8036 REV.1   CABLE LENGTH        : 10050M
C12  SHOT INTERVAL   : 37.5M              CABLE DEPTH         : 8M
C13  REC LENGTH/ST   : 14000MS/2MS        NO GROUPS/INTERVAL  : 804/12.5M
C14  HI FILTER       : 206HZ 276DB/OCT    LO FILTER           : 4.6HZ 6DB/OCT
C15  NAVIGATION      : SKYFIX XP          SPHEROID            : WGS 84
C16  START OF DATA DELAY : -50MS
C17  CENTRAL MERID.: 87 DEG E  - PROJECTION : UTM  45N - WORK DATUM : WGS 84
C18
C19  SOURCE  NUMBER 1
C20  CABLE NUMBER 1
C21  CHANNEL  NUMBER 1-804
C22
C23
C24                    SEG-Y FORMAT DESCRIPTION
C25  BYTES:  ITEM:            BYTES:  ITEM:            BYTES:   ITEM:
C26    9- 12 FIELD REC. NB      13- 16  FIELD CHANNEL     17- 20 SHOT POINT NB
C27   21- 24 SHOT POINT NB      37- 40  OFFSET            73- 76 SOURCE COORD. X
C28   77- 80 SOURCE COORD. Y    81- 84  GROUP COORD. X    85- 88 GROUP COORD. Y
C29  157-158 YEAR DATA REC.    159-160  DAY OF YEAR      161-162 HOUR OF DAY
C30  163-164 MINUTE OF DAY     165-166  SECOND OF DAY
C31  237-240 TIDAL STATIC CORRECTION IN MICROSECONDS
C32
C33  PROCESSING SEQUENCE:
C34  1) SEGD REFORMAT              3) START OF DATA DELAY CORRECTION -164MS
C35  3) NAVIGATION ASSIGNMENT      4) OUTPUT TO SEGY
C36
C37  POLARITY: AS RECORDED
C38  A NEGATIVE SAMPLE CORRESPONDS TO AN INCREASE IN ACOUSTIC IMPEDANCE
C39  END EBCDIC
C40

BINARY HEADER
-------------

Job ID:         0  Line Number:       103  Reel Number:         0
Traces          804 (Data)          0  (Auxiliary)
Sample Interval    2000 (Reel)      0  (Original) (in microsecs)
```

Figure 10 EBCIDIC Header SEGY data

Navigation Maps

Navigation Maps ensure that all the lines for a particular dataset have been transcribed and there are no missing lines or gaps. These navigation maps are generated from the actual trace data values and then compared with the available Navigation Maps received from the Acquisition Contractors.

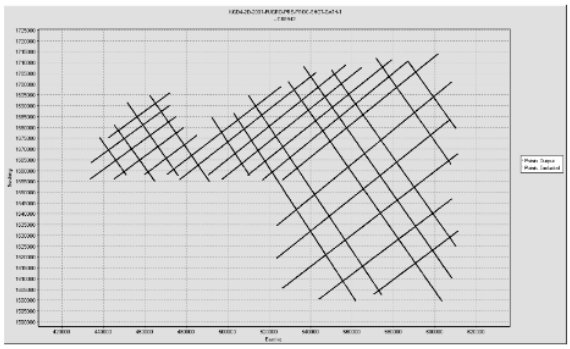


Figure 11 Navigation Map for 2D SEGY data

Data Readability & Ensuring Correctness of transcribed Output media

The transcribed output media is QC checked for readability through industry standard applications like Landmark's Promax and CGG Veritas Geocluster.

Conclusions

i. A Standard Operating Procedure (SOP) and robust multi-level QA/QC procedure is developed, tested and implemented.
ii. Greater than One Petabyte of data is transcribed without any errors or data loss.
iii. The entire SOP is being measured for throughput rate with timely and high quality results.
iv. The measured cycle times are suitable to be a benchmark for seismic data tape transcription.
v. This SOP and Workflow is proposed and recommended to be adopted as an E&P Industry standard for transcription of Seismic data.

References

Internal Reports
Project Scope of Work

Acknowledgments

We would like to thank Reliance Industries Ltd. Petroleum Business (E&P) for their support and permission for publishing this work. We would also like to thank the entire Project team at Reliance and Samit Spectrum for their continued efforts in keeping up the quality and meeting timelines of the Project.