DH07

Robust Study Database Build: An Effective Start Towards Creation of Quality SDTMs

By: Onkar Kajarekar

Introduction

01 Overview of a clinical trial

02 Database Build

03 Database Build >> Transformation >> Stats Programming

04 Dirty data – Cause and Impact

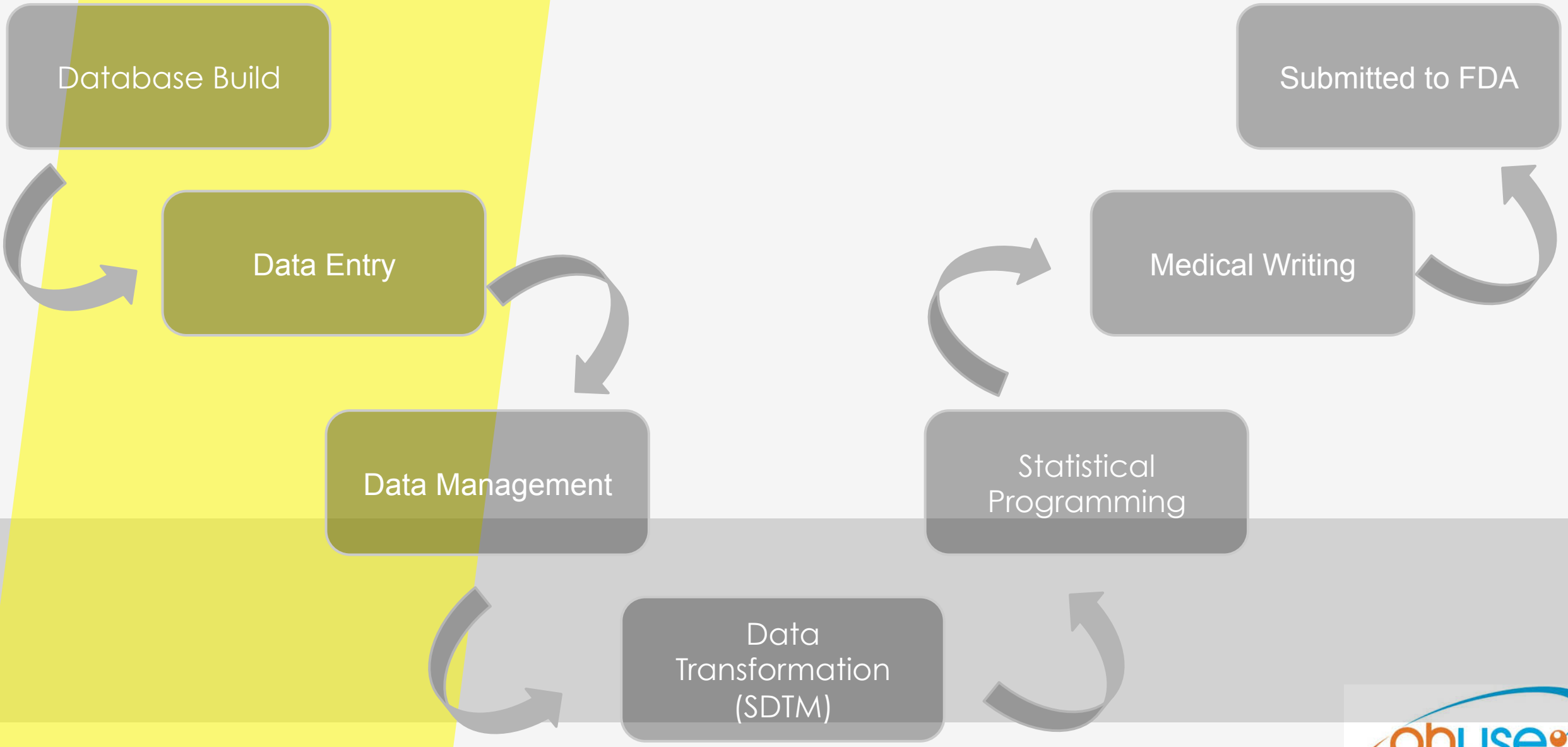
05 Example of Dirty data

06 Solution for Dirty data

07 Benefits

08 Conclusion

Overview of a Clinical Trial



Database Build

- Starting point for a clinical trial
- Database with GUI
- Enables electronic data entry
- Data Management Plan (DMP) – checks specifications
- Different platforms – RAVE, Oracle Clinical, INFORM and so on

Database Build >> Transformation

- Data Entry – Next step after Database build
- Data stored in SAS datasets (non-standard data)
- Data transformation (SDTM) – as per CDISC guidelines
- Aligned with FDA requirement

Table 1 – Disposition raw dataset

STUDY	SITE	PT	DISCRES	DISCDT
Study1	101	1001	SUBJECT WITHDREW	02OCT2010
Study1	101	1002	COMPLETED	03OCT2010

Table 2 – DS SDTM domain

USUBJID	DSCAT	DSTERM	DSDECOD	DSDTC
STUDY1-101-1001	DISPOSITION EVENT	SUBJECT WITHDREW	WITHDRAWAL BY SUBJECT	2010-10-02
STUDY1-101-1002	DISPOSITION EVENT	COMPLETED	COMPLETED	2010-10-03

Transformation ⟷ Stats Programming

Development of ADaMs using SDTMs as input

Development of TFLs using ADaMs as input

TFLs are used for further analysis by statisticians

Values displayed in the report purely depend on the data present in SDTMs and ADaMs

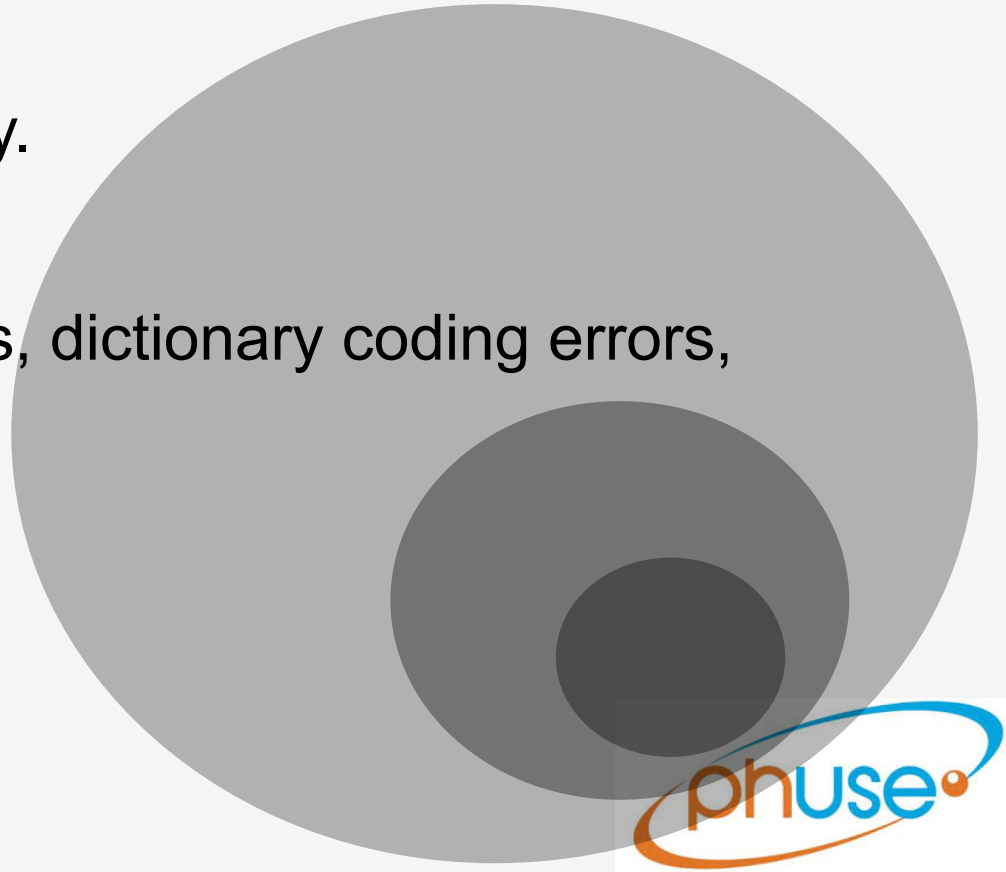
Figure1 – Summary of Disposition

Summary 1: Patient Disposition

Category	Total (N=2) n (%)
Completed Study	1 (50%)
Discontinued from Study	1 (50%)
Withdrawal by Subject	1 (50%)
Lost to Follow-up	0
Death	0

Dirty data – Cause

- Dirty data is nothing but unclean data which is not well organized, not properly cleaned and has no clinical significance.
- Errors inevitably occur due to incorrect data entry.
- Most common errors include typographical errors, dictionary coding errors, date mismatch errors and range errors



Dirty data – Impact

- Population of incorrect values in SDTMs and ADaMs
- Development of inaccurate statistical reports
- Impacts the quality of drug analysis
- Increased turnaround time for all the development operations
- Increased costs and inefficiencies
- Cost patient lives

Example of Dirty data

Consider a raw Adverse Event dataset and a raw subject disposition dataset. Below are the snapshots:

Table 3 – Adverse Event raw dataset

STUDY	SITE	PT	AERAW	AEOUT	AESEV	AESD	AEST	AEED	AEET
Study1	101	1001	VOMITING	FATAL	SEVERE	02OCT2010	9:30	02OCT2010	10:30

Table 4 – Disposition raw dataset

STUDY	SITE	PT	DISCRES	DISCDT
Study1	101	1001	SUBJECT WITHDREW	02OCT2010
Study1	101	1002	COMPLETED	03OCT2010

- In aforementioned scenario the discontinuation reason is populated as 'SUBJECT WITHDREW' because it has been captured that way in the raw dataset
- This can go unnoticed if there is no check implemented during the study build phase

Impact of Dirty data

Incorrect data entry will be reflected at SDTM level in AE and DS SDTM domain. Below are the snapshots of respective domains with dirty data:

Table 5 – AE SDTM domain

USUBJID	AETERM	AEOUT	AESEV	AESTDTC	AEENDTC
STUDY1-101-1001	VOMITING	FATAL	SEVERE	2010-10-02 T09:30	2010-10-02 T10:30

Table 6 – DS SDTM domain

USUBJID	DSCAT	DSTERM	DSDECOD	DSDTC
STUDY1-101-1001	DISPOSITION EVENT	SUBJECT WITHDREW	WITHDRAWAL BY SUBJECT	2010-10-02
STUDY1-101-1002	DISPOSITION EVENT	COMPLETED	COMPLETED	2010-10-03

Impact of Dirty data

This incorrect data would be reported at the Statistical programming level. Below are the snapshots of the tables and listings developed using the above SDTM data

Summary 1: Patient Disposition

Category	Total (N=2) n (%)
Completed Study	1 (50%)
Discontinued from Study	1 (50%)
Withdrawal by Subject	1 (50%)
Lost to Follow-up	0
Death	0

Listing 1: Listing of Adverse Events

Site/Subject	AE term	Outcome	Severity	Start Date	End Date
101/1001	Vomiting	Fatal	Severe	2010-10-02T09:30	2010-10-02T10:30

Robust Database Build – Proactive Solution for Dirty data

- Improvise the methodologies used to build the study database
- Proactive step taken at an initial stage of a clinical study
- Creation of Data Management Plan (DMP) for programming of checks to prevent the entry of dirty data
- Checks should cover maximum possible scenarios based on the protocol and CRF designing.

Robust Database Build – Proactive Solution for Dirty data

Once the check is implemented, the query will be automatically fired when inaccurate data is entered on any form as depicted in the below snapshot:

Figure 1 - Subject Disposition form

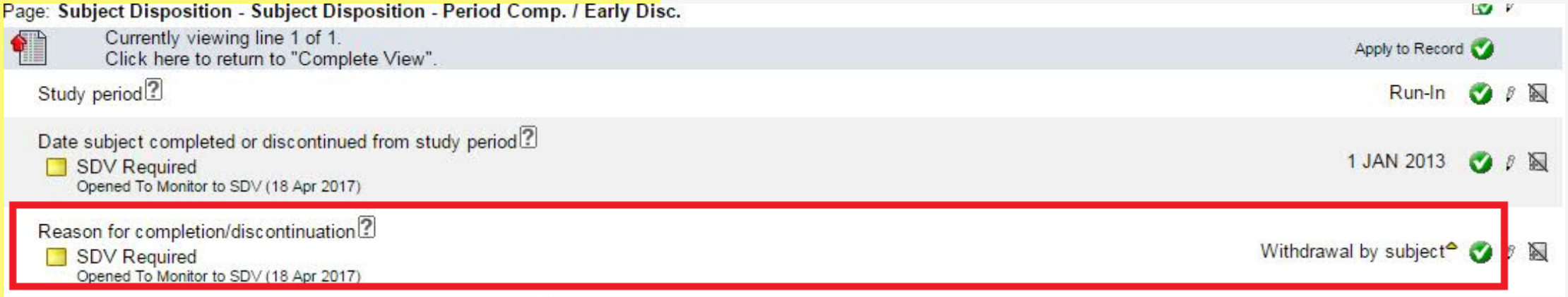
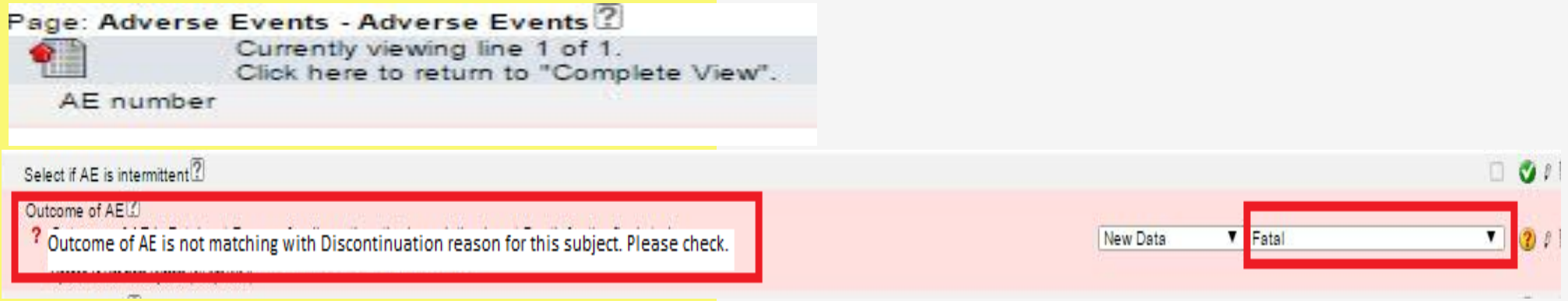


Figure 2 - Adverse Event form



Benefits

- Direct and positive impact on the subsequent phases of the study
- Enable the data entry team to query the issue to the corresponding site right in the beginning
- Reduction of data entry errors
- Seamless development of good quality SDTM domains and ADaMs
- Development of precise statistical reports
- Reduction in overall turnaround time and efforts
- On time FDA submissions

Benefits

- The raw disposition dataset and corresponding DS domain will look as follows after accurate data entry

Table 7 – Disposition raw dataset

STUDY	SITE	PT	DISCRES	DISCDT
Study1	101	1001	DEATH	02OCT2010
Study1	101	1002	COMPLETED	03OCT2010

Table 8 – DS SDTM domain

USUBJID	DSCAT	DSTERM	DSDECOD	DSDTC
STUDY1-101-1001	DISPOSITION EVENT	DEATH	DEATH	2010-10-02
STUDY1-101-1002	DISPOSITION EVENT	COMPLETED	COMPLETED	2010-10-03

Benefits

- Below are the tables created on the updated data in which the data is clearly ‘aligned’

Summary 2: Patient Disposition

Category	Total (N=2) n (%)
Completed Study	1 (50%)
Discontinued from Study	1 (50%)
Adverse Event	0
Lost to Follow-up	0
Death	1 (50%)

Listing 2: Listing of Adverse Events

Site/Subject	AE term	Outcome	Severity	Start Date	End Date
101/1001	Vomiting	Fatal	Severe	2010-10-02T09:30	2010-10-02T10:30

Conclusion

- Drug development process involves a huge financial investment by all the pharmaceutical companies globally
- Clinical trials are closely monitored in order to ensure that accurate results are obtained from both efficacy and financial standpoint
- Programming of key checks will certainly contribute significantly in achieving the enhanced data quality in efficient way
- Development of robust database will certainly prove an effective start to creation of good quality SDTMs and accurate statistical reports

Questions???

Name: Onkar Kajarekar

Organization: Tata Consultancy Services
Address: Empire Plaza, 4th floor,
Lal Bahadur Shastri Marg, Chandan Nagar,
Vikhroli West, Mumbai, Maharashtra 400083

Work Phone: +91-22-67786057
Email: onkar.kajarekar@tcs.com
Web: www.tcs.com

Thank You!