phuse·share
connect·share·advance

Single Day Event

Bridging the Gap – Transforming Unstructured eDT Data to Structured Data

Tanushree Srivastava,
Johnson & Johnson

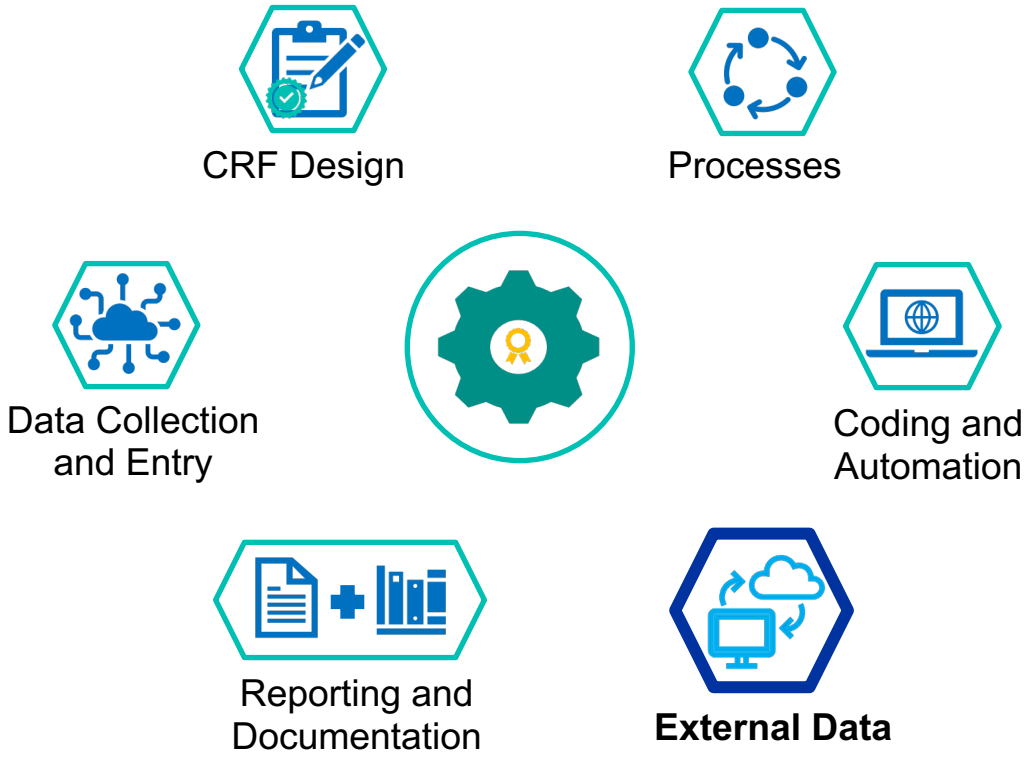
Disclaimer

This presentation is for informational purposes only and does not represent professional guidance or advice. Any views and opinions expressed during this presentation are those of the presenters and do not necessarily reflect the views or policies of Janssen Research & Development, LLC, or any other company in the Johnson & Johnson Family of Companies.

Agenda

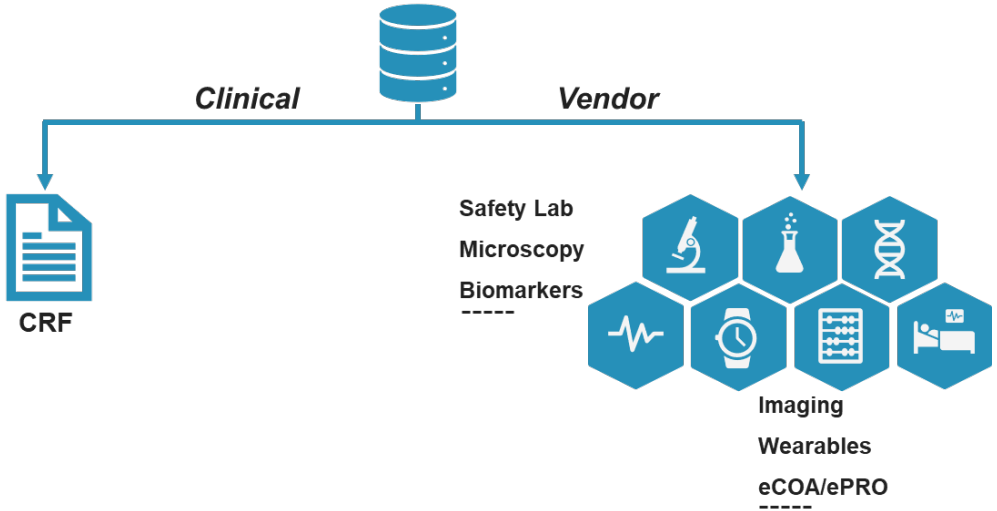
- Introduction
- Challenge
- Solution
- Process
- Benefits

Introduction -> Standardization



Introduction

- Clinical data originates from different sources with varying complexity
- Data collection includes receiving data from electronic CRF (and/or paper CRF) and external vendor data



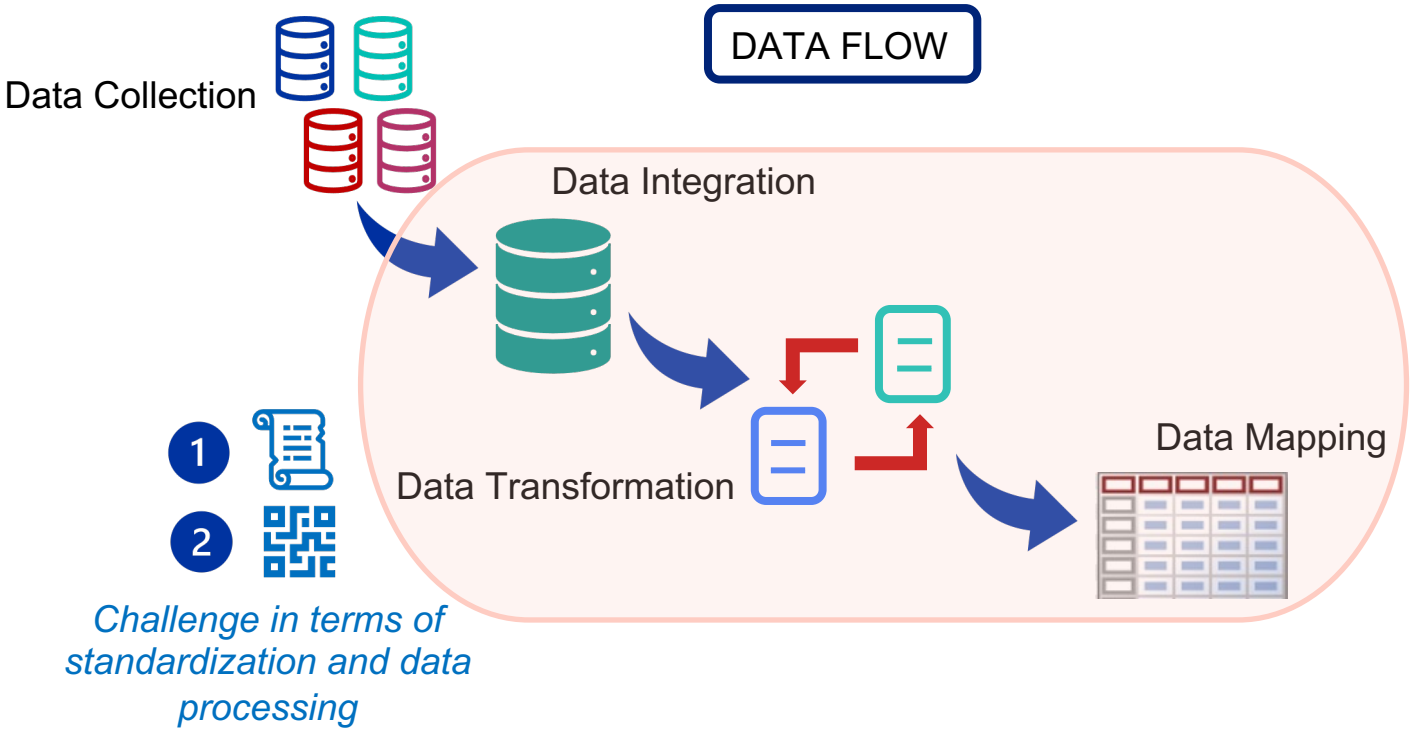
Introduction

- With the growth in the industry, volume of data is also increasing

→ Effective management of data

? Standardize external data processing across vendors and therapeutic area

Challenge


DATA FLOW
Data Collection
Data Integration
Data Transformation
Data Mapping
1
2
Challenge in terms of standardization and data processing

Challenge

Transfer variable name	Transfer variable label	Format	Length
Study_ID	Study ID	A/N	15
Screen_ID	Screen ID	A/N	30
Result_Name	Result Name	A	60
Result_Value_Literal	Result Value Literal	A/N	225
Results_Units	Result Units	A	20
Visit	Visit	A/N	22
Date_Collected	Date Collected	N	8
Time_Collected	Time Collected	N	5

Transfer variable name	Transfer variable label	Format	Length
STUDYID	Study Identifier	Char	40
SUBJID	Subject ID or Number	Char	8
LBTEST	Lab Test or Examination Name	Char	40
LBORRES	Result or Finding in Original Units	Char	200
LBORRESU	Original Units	Char	50
VISIT	Visit Name	Char	60
LBDTC	Date/Time of Specimen Collection	Char	19

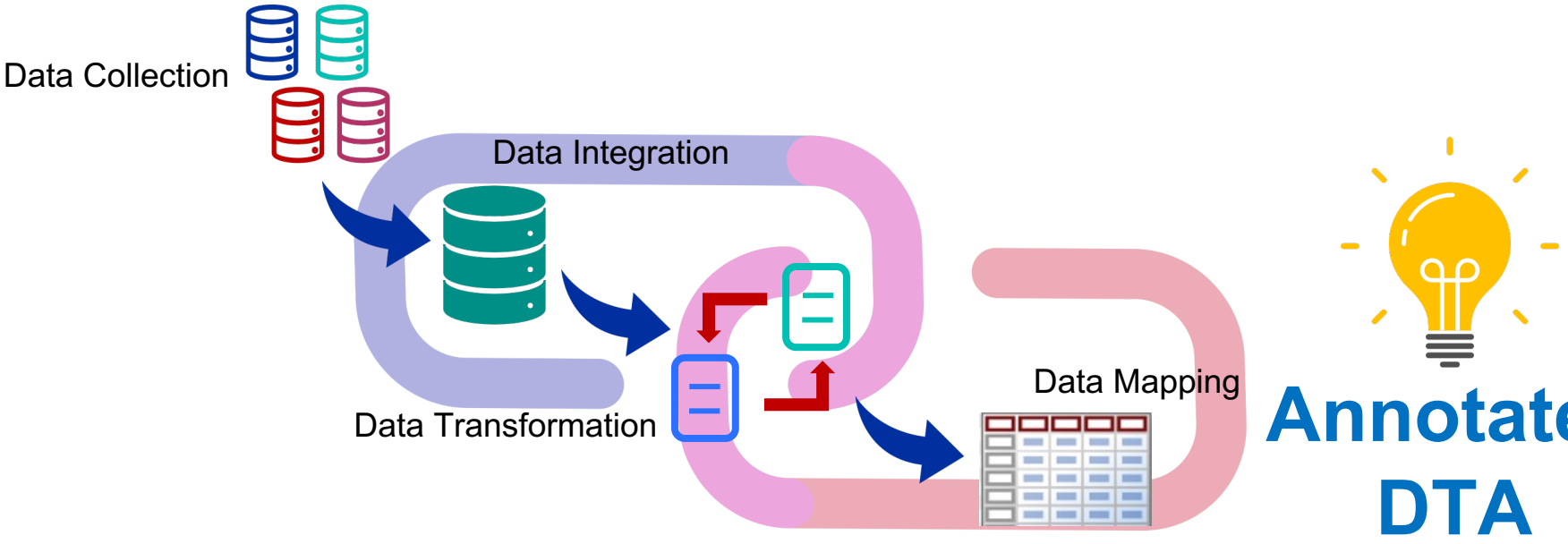
- Not all vendors deliver data that is aligned with CDISC SDTM standards
- Different transfer capabilities of vendors
- Different internal vendor standards

Challenge

Transfer variable name	Transfer variable label	Format	Length	Description of transfer variable use
Study_ID	Study ID	A/N	15	Study identifier
Date_of_service	Date Of Service	N	8	Format: MM/DD/YYYY Date Logged into Lab System
Date_Collected	Date Collected	N	8	Format: YYYY-MM-DD Date the specimen was collected
Time_Collected	Time Collected	N	5	Format: HH:MM Time the specimen was collected
Screen_ID	Screen ID	A/N	30	Quest OE field Patient ID will be used
Visit	Visit	A/N	22	Quest OE Lab reference ID
Order_Name	Order Name	A	60	Name of test requested by client
Result_Name	Result Name	A	60	Local Result Name
Result_Value_Literal	Result Value Literal	A/N	225	Contains both alpha and numeric results
Results_Units	Result Units	A	20	Units of measure associated with result report (Not always reported)

- Internal post-processing
- Inconsistent data mapping and alignment with standards
- Lack of documentation

Solution

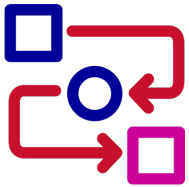

Data Collection
Data Integration
Data Transformation
Data Mapping
Annotated
DTA

Annotated DTA

- Macro enabled excel worksheet
- Standalone document
- Not part of submission

- Data delivery expectations are imported from tsDTA
- Documents the mapping specifications of source to target
- Automated Conversion - incoming data to SAS dataset
- Automated QC - checks to accept/reject the transfer

Process -> Automated Conversion



Link to tsDTA		RAW DATA STRUCTURE (as per tsDTA)				
tsDTA version	Transfer Metadata Tab Name	Transfer file order	Transfer variable name	Transfer variable label	Format	Anticipated Maximum Length

Import tsDTA

Link to tsDTA		RAW DATA STRUCTURE (as per tsDTA)				
tsDTA version	Transfer Metadata Tab Name	Transfer file order	Transfer variable name	Transfer variable label	Format	Anticipated Maximum Length
1	Transfer Metadata	1	STUDYID	Study Identifier	Char	40
1	Transfer Metadata	2	DOMAIN	Domain Abbreviation	CHAR	2
1	Transfer Metadata	3	LBTEST	Lab Test or Examination Name	CHAR	40
1	Transfer Metadata	4	VISIT	Visit Name	CHAR	60
1	Transfer Metadata	5	SUBJID	Subject ID or Number	CHAR	40

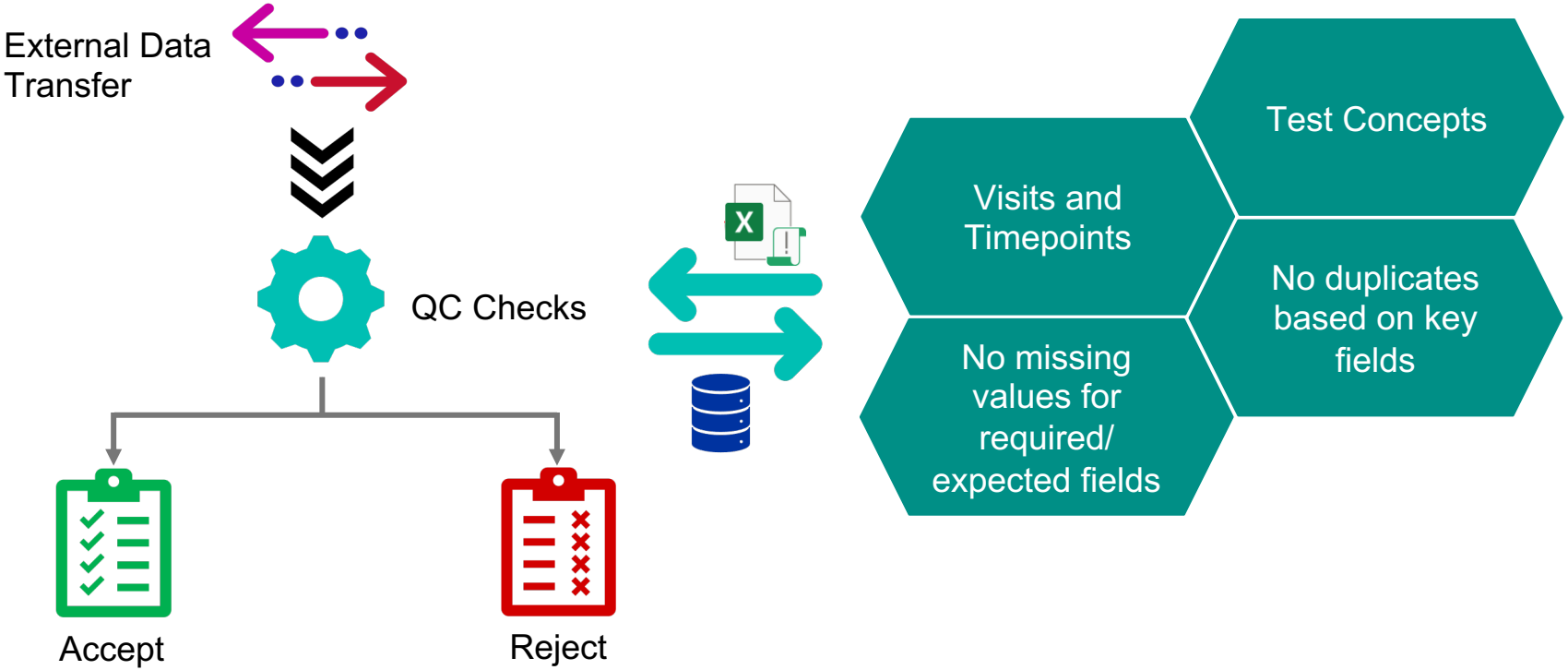
SAS CONVERSION			
SAS VARIABLE NAME	SAS LABEL	SAS TYPE	SAS LENGTH
STUDYID	Study Identifier	character	40
DOMAIN	Domain Abbreviation	character	2
LBTEST	Lab Test or Examination Name	character	40
VISIT	Visit Name	character	60
SUBJID	Subject ID or Number	character	30

IRM SDTM

Link to tsDTA		RAW DATA STRUCTURE (as per tsDTA)					INTERNAL REVIEW MODEL ANNOTATION [1,2]				SDTM ANNOTATION			
tsDTA version	Transfer Metadata Tab Name	Transfer file order	Transfer variable name	Transfer variable label	Format	Anticipated Maximum Length	TARGET Domain	TARGET Variable	TARGET Variable Label	DRM Mapping Guidance	TARGET Domain	TARGET Variable	TARGET Variable Label	SDTM Mapping Guidance
1	Transfer Metadata	1	STUDYID	Study Identifier	Char	40	DRM_LB	STUDYID_EDT_IRM STUDYID RSTUDYID	Source EDT Study Identifier Study Identifier Raw Study Identifier	Copy vendor value Recode value Copy vendor value	LB	STUDYID	Study Identifier	Copy vendor value
1	Transfer Metadata	2	DOMAIN	Domain Abbreviation	CHAR	2	DRM_LB	DOMAIN_EDT_IRM	Source EDT Domain Abbreviation	Copy vendor value	LB	DOMAIN	Domain Abbreviation	Copy vendor value
1	Transfer Metadata	3	LBTEST	Lab Test or Examination Name	CHAR	40	DRM_LB	LBTEST_EDT_IRM LBTEST	Source EDT Lab Test or Examination Name Lab Test or Examination Name	Copy vendor value See Test Concepts	LB	LBTEST	Lab Test or Examination Name	Copy vendor value See Test Concepts
1	Transfer Metadata	4	VISIT	Visit Name	CHAR	60	DRM_LB	VISIT_EDT_IRM VISIT RVISIT	Source EDT Visit Name Visit Name Raw Visit	Copy vendor value See Visits and Timepoints Copy vendor value	LB	VISIT	Visit Name	Copy vendor value See Visits and Timepoints
1	Transfer Metadata	5	SUBJID	Subject ID or Number	CHAR	40	DRM_LB	RSUBJECT	Raw Subject Identifier	Copy vendor value	LB	SUBJID	Subject ID or Number	Copy vendor value

1. Proprietary, CDASH/SDTM-hybrid data model to expedite clinical data review, Paper SI25. L. Gijsbers, T. Van der Spiegel. s.l. : PhUSE, 2018.
2. A proprietary, CDASH/SDTM-hybrid data model to expedite clinical data review. M.A. Pordhomme, L. Gijsbers, T. Van der Spiegel. s.l. : CDISC, 2019.

Process -> Automated QC



Automated QC - Example

LBTESTCD	LBTEST	LBCAT	LBORRESU	LBNAM	LBSPEC
IFNG	Interferon Gamma	IMMUNOLOGY	pg/mL	LAB	SERUM
INTLK10	Interleukin 10	IMMUNOLOGY	pg/mL	LAB	SERUM
IL12P70	Interleukin 12 Subunit P70	IMMUNOLOGY	pg/mL	LAB	SERUM
INTLK13	Interleukin 13	IMMUNOLOGY	pg/mL	LAB	SERUM

Visit	Visit Number	Timepoint	Timepoint number
STEP-UP DOSE 1 DAY 1	201001	PRE-DOSE	-0.001
STEP-UP DOSE 1 DAY 2	201002	24H	24
DOSE 1 DAY 1	301001	PRE-DOSE	-0.001
DOSE 1 DAY 2	301002	24H	24
DOSE 1 DAY 2	301002	36H	36
DOSE 1 DAY 3	301003	72H	72

```
LBTESTCD|LBTEST|LBCAT|LBORRESU|LBNAM|LBSPEC|VISIT
IFNGA|Interferon Gamma|IMMUNOLOGY|pg/mL|LAB|SERUM|DOSE 2 DAY 1
ALB|Albumin|IMMUNOLOGY|pg/mL|LAB|SERUM|DOSE 1 DAY 1
```

NOTE CHECK ID: **[_adf_002_check_adta_vs_01]** EXECUTED WITH RESULT **FAIL** FOR JOB ID: [20240112060702] INVOLVING: lab_lb
Please investigate: [The issues are listed below.
For test concept L BCTKTEST the values (separated by pipe) do not match for columns LBTESTCD LBTEST LBCAT LBORRESU LBNAM LBSPEC LBMETHOD CODE: lab_lb
(row 1 of 2) ALB|Albumin|IMMUNOLOGY|pg/mL| LAB |SERUM|10235
(row 2 of 2) IFNGA|Interferon Gamma|IMMUNOLOGY|p g/mL| LAB |SERUM|10235]

NOTE CHECK ID: **[_adf_002_check_adta_vs_01]** EXECUTED WITH RESULT **FAIL** FOR JOB ID: [20240112060702] INVOLVING:
Please investigate: [The issues are listed below.
The value - DOSE 2 DAY 1 is not expected for the variable VISIT in lab_lb dataset. The expected values are - DOSE 1 DAY 1,DOSE 1 DAY 2,DOSE 1 DAY 3,DOSE 1 DAY 7,STEP-UP DOSE 1 DAY 1,STEP-UP DOSE 1 DAY 2,STEP-UP DOSE y DAY 1,STEP-UP DOSE y DAY 2,UNSCHEDULED.]

Benefits

- Proper and transparent documentation of raw to target mapping of external data
- Consistent way of documenting raw to target mapping of external data
  - Facilitates End-to-End standardization
  - Alignment with standards
  - Auditability, traceability and oversight

- Excel approach allows first steps towards machine readable validation
- Automated conversion and downstream utilization
- Reduced mapping time

Thank you!

The Global Healthcare
Data Science Community

Contact Channels
UK +44 1843 609600
office@phuse.global
phuse.global

Social Media
@phusetwitta
/phusebook
/phusetube
/company/phuse

Presenter Contact Detail:
tsrivas2@its.jnj.com