PharmaSUG 2018 - Paper SS-05

Challenges for Implementing Legacy Data Conversion Plan (LDCP) in Electronic Data Submission

Shengfeng (Philip) Ho, Xiaojia (Jessie) Wang, Rundo International Pharmaceutical Research & Development Co.,LTD

ABSTRACT

The development of a Legacy Data Conversion Plan (LDCP) is identified as a need in recent released Study Data Technical Conformance Guide in August 2017 which associated with FDA Guidance on Providing Regulatory Submissions in Electronic Format – Standardized Study Data (Final FDA Guidance December 2014). Unlike Study Data Standardization Plan, LDCP serves for legacy study and assists regulatory reviewers in understanding your data conversion process and possible traceability issue. This paper presents the challenges for implementing LDCP in legacy study for ISS/ISE submission. The case study and examples shown in this paper are our working experiences while leading a group of programmers in the effort to convert data for 9 sub-studies from legacy format to industry-standard.

INTRODUCTION

As FDA announces the deadline, December 17, 2016 and series of binding guidance, that require clinical and non-clinical study submission to use standards supported in the FDA Data Standards Catalog [1]. There is also a need for legacy study data[1] to convert per the standards in the catalog. In study data technical conformance guide (SDTCG) v4.0, the document clarifies related guidance on converting legacy data to standard format and provides possible limitations of keeping traceability. Also, Legacy Data Conversion Plan (LDCP) firstly introduced in the document to serves as an in-front plan for regulatory reviewers to easily understand your conversion details and how you maintain the traceability, which is going to be combined with clinical Study Data Reviewer's Guide as a whole document. Legacy study data conversion should have following characteristics:

a. Map every data element as originally collected
b. May not be possible to represent a collected data element as a standardized data element
c. Omitted data should be apparent on the annotated CRF and described in the reviewer's guide

As we may know, most legacy studies may not have well-documented or conducted before CDISC standards released. Once the legacy data structure doesn't follow CDASH standard to design or compatible with other standards, it will increase difficulties to perform data conversion to CDISC standard format.

Moreover, in the recent survey about the use of CDISC standards in Pharmaceutical industry [3], it shows that over half (52%) of the respondents[2] viewed the difficulty in building governance processes on implementing standards as the primary organizational challenge, 33% respondents still used spreadsheets to manage standards and study specifications, and 44 % respondents cited lack of internal CDISC knowledge and experience as a barrier to standards adoption. So, a professional team and robust processes/tools are the keys to drive how fast you can prepare study submission package.

Therefore, the challenges you can foresee for legacy study data conversion are from its naive deficiency of data acquisition design and limitation of utilizing automated metadata tool to complete the data conversion. In following sections, we describe the LDCP implementation by using a case study and

[1] Legacy study data are study data in a non-standardized format, not supported by FDA and not ever listed in the Catalog.
[2] Respondents were high level executives with extensive experience from a range of sub-sectors including pharmaceutical, biotech, med device manufacturers, academic research centers, contract research organizations.

discuss some outstanding issues with regards to traceability and legacy data deficiency. Finally, we wrap up with suggestions to perform legacy data conversion in a more efficient way.

THE CASE STUDY & LEGACY DATA CONVERSION APPROACH

The case study consisted of 9 independent sub-studies which conducted from 1996 to 2012 and belong to CNS/Psychiatry therapeutic area. The locked data were used to convert to SDTM and ADaM compliant datasets in sequence. The conversion was led by 9 programming teams with one programmer lead's coordination.

During development of the mapping specification from legacy data to SDTM, CDISC Controlled Terminology managed in spreadsheet was applied where applicable. Firstly, SDTM annotated CRF was created to summarize how many SDTM domains were required. After authoring of a mapping specification and programming of the SDTM SAS datasets, the Pinnacle21 validator was run to check compliance to SDTM v1.4/ SDTMIG v3.2.

The QC step was performed where the datasets were double-programmed by an independent QC programmer using the same mapping specification as the reference. Any fall outs were recorded in the unique tracking sheet and returned to the SDTM programmer for updates. After confirmation that updates were applied properly, the SDTM data was considered complete then SDTM define.xml created by SAS with specification-driven approach. The ADaM datasets were derived from the SDTM datasets by ADaM v2.1/ADaMIG v1.0 under the similar process.

CSR needs to take as input to decide missing data imputation or data mapping issues throughout the conversion. Finally, all ADaM datasets were pooled together for ISS/ISE analysis. Figure 1 presents the legacy data conversion flow in the case study.

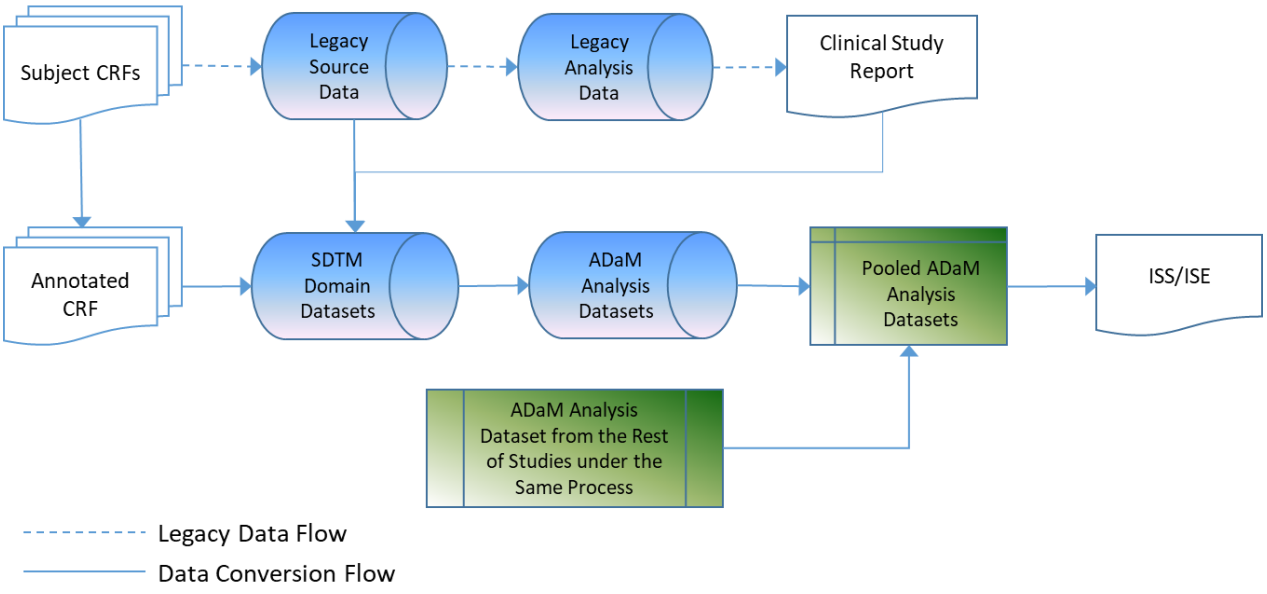


Figure 1. The Legacy Data Conversion Flow of Case Study

The conversion approach for the case study can also be briefly summarized by Clinical Data Standards Capability Maturity Model [3] as below three dimensions:

1. CDISC Standard – Use of CDISC Standards for Compliance in Regulatory Submissions
2. Standards Metadata Management and Use – Siloed, Manual Management of Spreadsheet based Metadata; Limited Metadata-driven Processing
3. Standards governance – Siloed Standards Governance with Limited Dedicated Staff

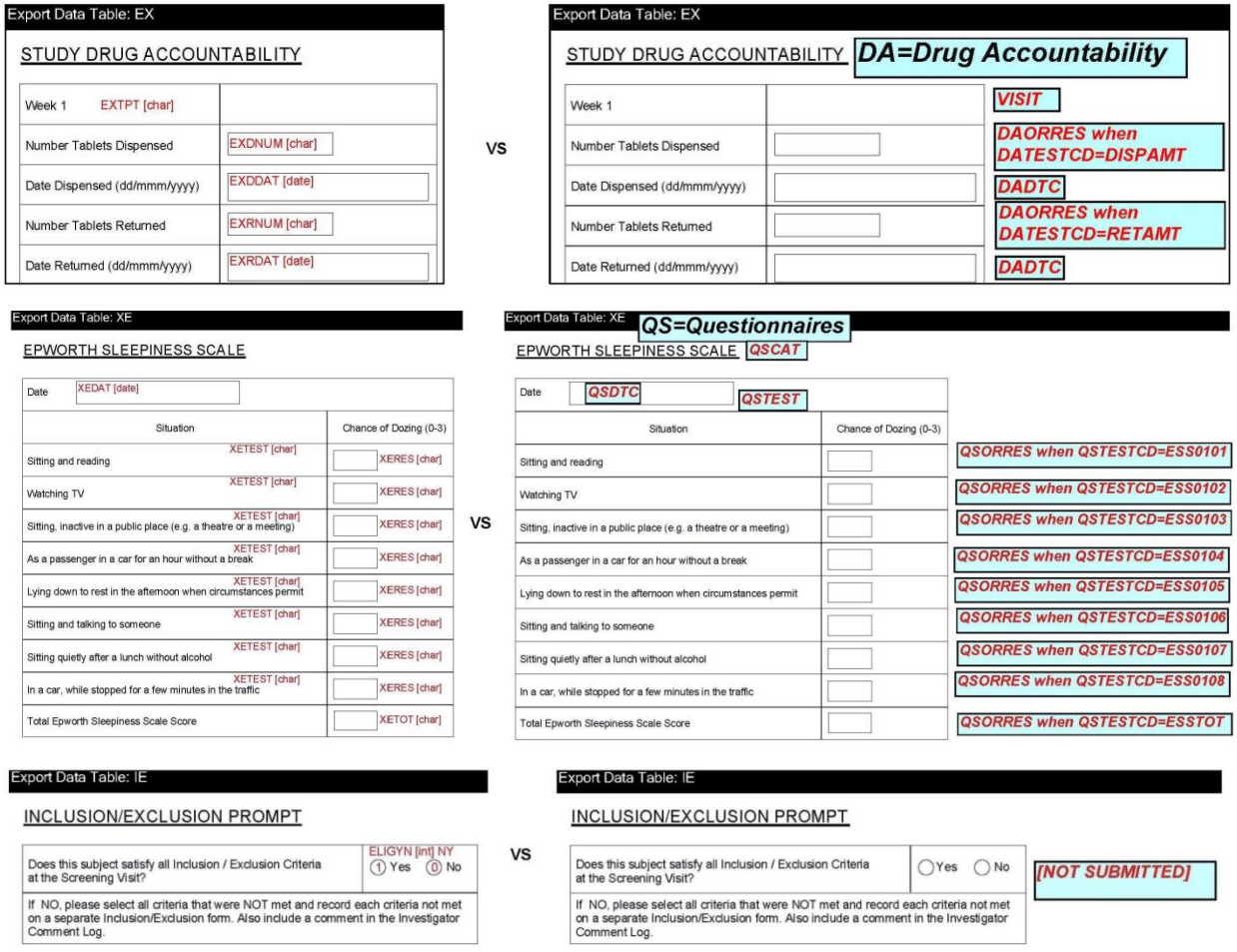
OUTSTANDING ISSUE SUMMARY & WORKAROUND

From annotating CRF through creating ADaM analysis dataset, some challenging issues obstacle us to rebuild the traceability. In the section, we focus on traceability limitations listed in table 5 of SDTCG and give some examples to demonstrate our workaround. In addition, other examples are put together in last limitation, called naïve data deficiency from legacy study.

1. LIMITED ABILITY TO DETERMINE LOCATION OF COLLECTED CRF VARIABLES IN THE CONVERTED SDTM DATA UNLESS THE LEGACY ACRF IS RE-ANNOTATED

1) Legacy Annotation Misled to SDTM Annotation

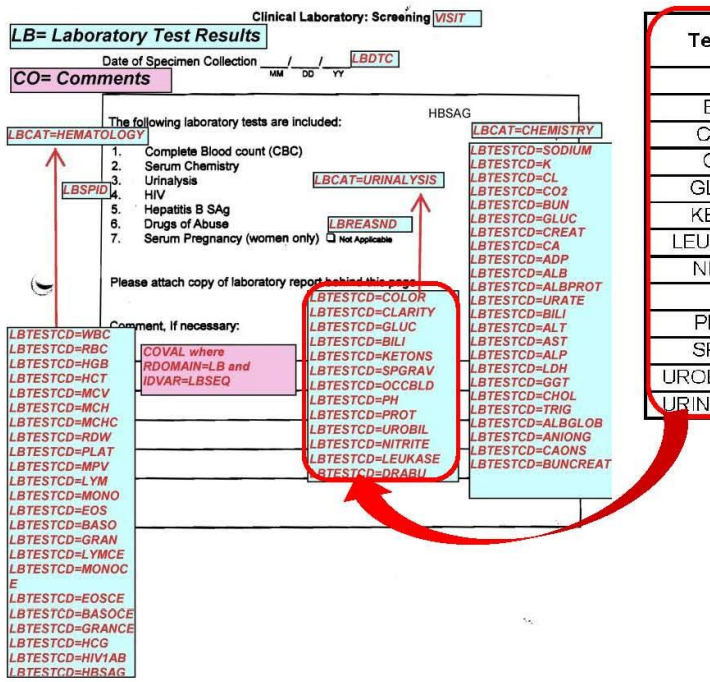
Legacy annotation in some forms was not compliant with SDTM. After discussed with sponsor, CRF should re-annotate per SDTM/SDTMIG and metadata submission guide. Display 1 compares the annotations between legacy and SDTM.



Display 1. Comparison between Legacy and SDTM Annotation

2) Too Simple CRF Design to Annotate

In a sub-study, laboratory test assessed by central laboratory vendor and the CRF just listed all categories of laboratory tests. For SDTM aCRF, we re-annotated all LBTESTCDs in the laboratory test form by using the test name in vendor dataset. Display 2 takes Urine Analysis and external dataset as example to illustrate the re-annotations.



Test Name	Test Name coded	Units	Range
BILI	1	NS	Negative
BLOOD	2	NS	Negative
CLARITY	3	NS	Not Specified
COLOR	4	NS	Not Specified
GLUCOSE	5	NS	Negative
KETONES	6	NS	Negative
LEUKOCYTES	7	NS	Negative
NITRITES	8	NS	Negative
PH	9	NS	4-8
PROTEIN	10	MG/DL	Negative
SPECIFIC	11	NS	1.005-1.035
UROBILINOGEN	12	EU/DL	0.1-1
URINE OPIATES	13	NG/ML	

Display 2. The Clinical Laboratory Case Report Form and External Urine Analysis Data

2. LIMITED TRACEABLE PATH FROM SDTM TO THE LEGACY ANALYSIS DATA

Not applicable to the case study since legacy analysis datasets were not used to ISS/ISE.

3. LIMITED ABILITY TO REPLICATE/CONFIRM LEGACY ANALYSIS DATASETS (I.E., ANALYSIS VARIABLE IMPUTATION OR DERIVED VARIABLES) USING SDTM DATASETS

Not applicable to the case study since legacy analysis datasets were not used to ISS/ISE.

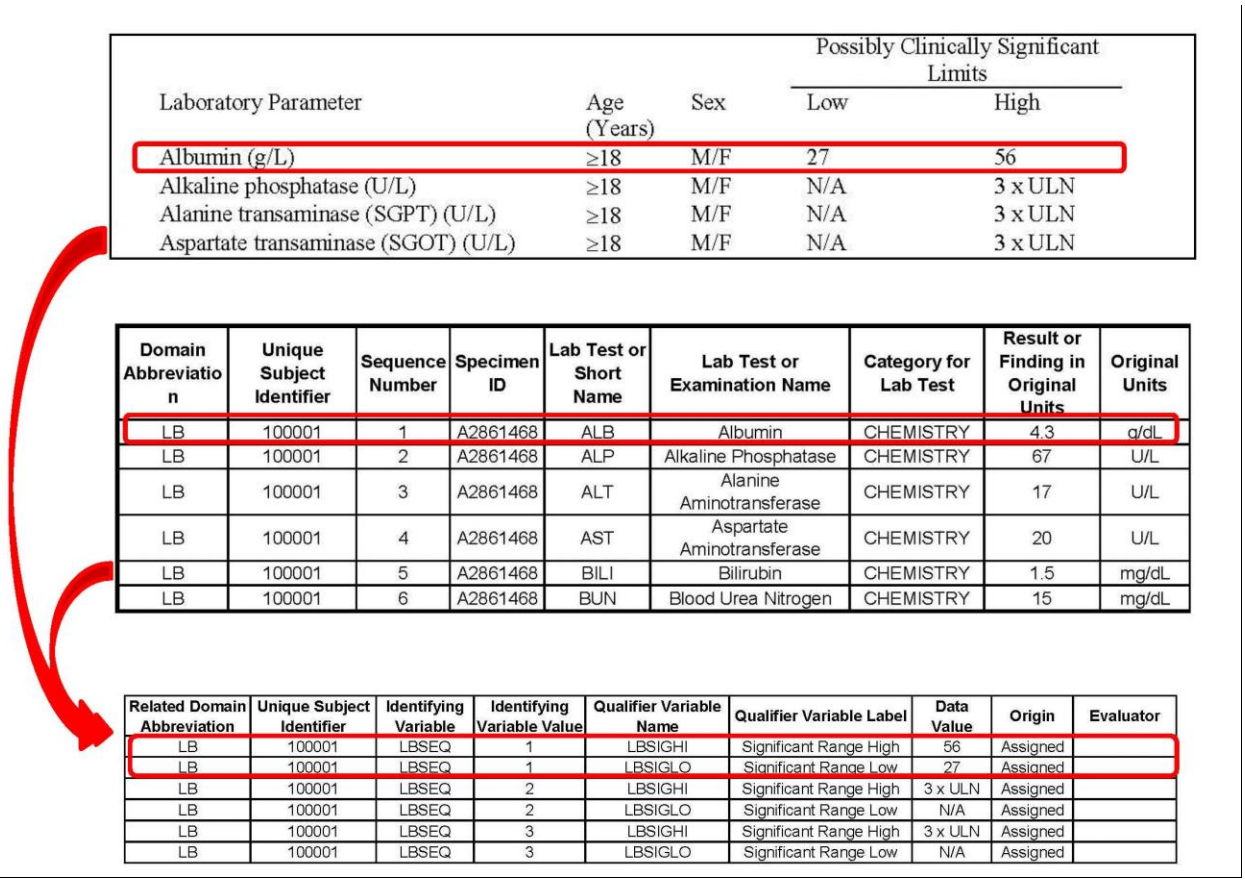
4. LIMITED ABILITY TO CONFIRM DERIVATION OF INTERMEDIATE ANALYSIS DATASETS OR CUSTOM DOMAINS

Not applicable to the case study since intermediate analysis datasets were not used to ISS/ISE.

5. LIMITED TRACEABLE PATH FROM ADAM TO THE TABLES, FIGURES AND THE CSR

1) Laboratory Clinically Significant Reference Range Referred to CSR

The “Possibly Clinically Significant Limits” in CSR was determined as source data to map into SUPPLB domain. This is because the source document of the reference range had been lost. The Display 3 takes Albumin test as example to show the traceability of clinical significant limits from CSR to SUPPLB domain.



Laboratory Parameter	Age (Years)	Sex	Possibly Clinically Significant Limits Low	High
Albumin (g/L)	≥18	M/F	27	56
Alkaline phosphatase (U/L)	≥18	M/F	N/A	3 x ULN
Alanine transaminase (SGPT) (U/L)	≥18	M/F	N/A	3 x ULN
Aspartate transaminase (SGOT) (U/L)	≥18	M/F	N/A	3 x ULN

Domain Abbreviation	Unique Subject Identifier	Sequence Number	Specimen ID	Lab Test or Short Name	Lab Test or Examination Name	Category for Lab Test	Result or Finding in Original Units	Original Units
LB	100001	1	A2861468	ALB	Albumin	CHEMISTRY	4.3	g/dL
LB	100001	2	A2861468	ALP	Alkaline Phosphatase	CHEMISTRY	67	U/L
LB	100001	3	A2861468	ALT	Alanine Aminotransferase	CHEMISTRY	17	U/L
LB	100001	4	A2861468	AST	Aspartate Aminotransferase	CHEMISTRY	20	U/L
LB	100001	5	A2861468	BILI	Bilirubin	CHEMISTRY	1.5	mg/dL
LB	100001	6	A2861468	BUN	Blood Urea Nitrogen	CHEMISTRY	15	mg/dL

Related Domain Abbreviation	Unique Subject Identifier	Identifying Variable	Identifying Variable Value	Qualifier Variable Name	Qualifier Variable Label	Data Value	Origin	Evaluator
LB	100001	LBSEQ	1	LBSIGHI	Significant Range High	56	Assigned	
LB	100001	LBSEQ	1	LBSIGLO	Significant Range Low	27	Assigned	
LB	100001	LBSEQ	2	LBSIGHI	Significant Range High	3 x ULN	Assigned	
LB	100001	LBSEQ	2	LBSIGLO	Significant Range Low	N/A	Assigned	
LB	100001	LBSEQ	3	LBSIGHI	Significant Range High	3 x ULN	Assigned	
LB	100001	LBSEQ	3	LBSIGLO	Significant Range Low	N/A	Assigned	

Display 3. The Traceability of Clinical Significant Limits for Laboratory Tests from CSR to SUPPLB

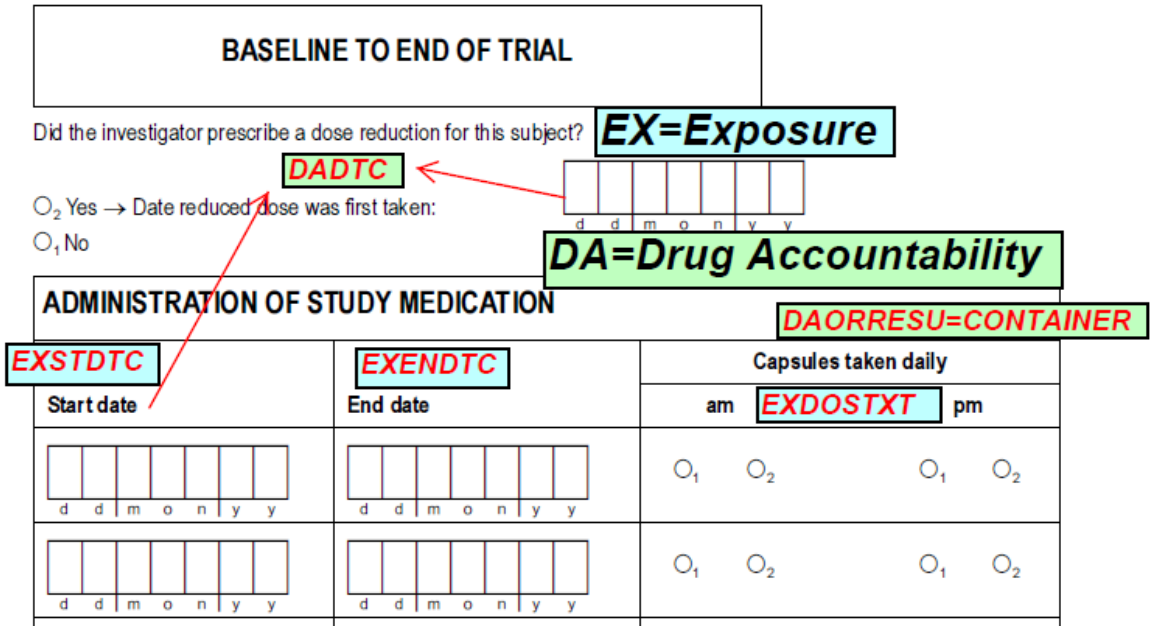
2) EXDOSTXT Mapping Design Due to The Derivation Requirement in ADEX

In one sub-study design, two arms would be administered in escalating doses up to the specified target dose, active reference compound and placebo would be given at constant dose. During each treatment day, subjects were supposed to take 4 capsules every day (two capsules in the morning and two capsules in the evening no matter of study medication or placebo or active reference compound, but the capsules were designed to have different strength (0 or 100mg).

For example, if a subject was randomly assigned to the 400 mg group, then from the first day to the next day, the first capsule he took every morning was a capsule containing an effective dose of 100 mg. He took capsules on the first and second days. The effective dose content was 100 mg in total. From the third day to the fourth day, he took the first capsule every morning and afternoon was a capsule containing an effective dose of 100 mg. The effective dose of his capsules on the third and fourth days was 200 mg in total. From the fifth day to the seventh day, the two capsules he took every morning and the first capsule took in the afternoon contained an effective dose of 100 mg capsules. The effective dose of capsules he took on the fifth to the seventh days was 300 mg in total. From the eighth day to the fourteenth day, two capsules he took every morning and two capsules took in the afternoon contained effective doses of 100 mg capsules. He had 400 mg total effective dose of capsules on the eighth and fourteenth days. After the fifteenth day, if the subject did not show symptoms of discomfort, the capsules were always taken from the eighth day to the fourteenth day until the end of the study. The scheduled active dosage for study drug administration and the administration of study medication CRF is shown in Table 1 and Display 4, respectively, for illustration.

Day	200 mg/day(max.)				400 mg/day(max.)				Active reference compound 20 mg/day (max.)	Placebo
	AM1	AM2	PM1	PM2	AM1	AM2	PM1	PM2		
Day 1-2	100mg	0mg	0mg	0mg	100mg	0mg	0mg	0mg	20 mg	Placebo
Day 3-4	100mg	0mg	0mg	100mg	100mg	0mg	100mg	0mg	20 mg	Placebo
Day 5-7	100mg	0mg	0mg	100mg	100mg	100mg	100mg	0mg	20 mg	Placebo
Day 8-14	100mg	0mg	0mg	100mg	100mg	100mg	100mg	100mg	20 mg	Placebo
Day 15-21	100mg	0mg	0mg	100mg	100mg	100mg	100mg	100mg	20 mg	Placebo
Day 22	100mg	0mg	0mg	100mg	100mg	100mg	100mg	100mg	20 mg	Placebo
Day 23 to end	100mg	0mg	0mg	100mg	100mg	100mg	100mg	100mg	20 mg	Placebo

Table 1. The Scheduled Active Dosage for Study Drug Administration



Display 4. Administration of Study Medication

Because capsule numbers and dosage information would be used to derive compliance in ADEX dataset (shown in Table 2), after discussed with study statistician, we decided to concatenate all required information, including AM/PM, the order of administration, number of capsule taken, and dosage in EXDOSTXT variable in EX domain. Table 3 shows one subject's completed exposure records in final EX domain.

Unique Subject Identifier	Source Domain	Source Sequence	Planned Treatment	Actual Treatment	Analysis Dose per Administratio	Dose Description Dose, mg/day	Total Daily	Start Date/Time of Treatment	End Date/Time of Treatment
100008	EX	1	Study_M 400 mg/day	Study_M 400 mg/day	100	AM1_1_100	100	22-Aug-03	23-Aug-03
100008	EX	2	Study_M 400 mg/day	Study_M 400 mg/day	0	AM2_1_0	100	22-Aug-03	23-Aug-03
100008	EX	3	Study_M 400 mg/day	Study_M 400 mg/day	0	PM1_1_0	100	22-Aug-03	23-Aug-03
100008	EX	4	Study_M 400 mg/day	Study_M 400 mg/day	0	PM2_1_0	100	22-Aug-03	23-Aug-03
100008	EX	5	Study_M 400 mg/day	Study_M 400 mg/day	100	AM1_1_100	200	24-Aug-03	25-Aug-03
100008	EX	6	Study_M 400 mg/day	Study_M 400 mg/day	0	AM2_1_0	200	24-Aug-03	25-Aug-03
100008	EX	7	Study_M 400 mg/day	Study_M 400 mg/day	100	PM1_1_100	200	24-Aug-03	25-Aug-03
100008	EX	8	Study_M 400 mg/day	Study_M 400 mg/day	0	PM2_1_0	200	24-Aug-03	25-Aug-03
100008	EX	9	Study_M 400 mg/day	Study_M 400 mg/day	100	AM1_1_100	300	26-Aug-03	28-Aug-03
100008	EX	10	Study_M 400 mg/day	Study_M 400 mg/day	100	AM2_1_100	300	26-Aug-03	28-Aug-03
100008	EX	11	Study_M 400 mg/day	Study_M 400 mg/day	100	PM1_1_100	300	26-Aug-03	28-Aug-03
100008	EX	12	Study_M 400 mg/day	Study_M 400 mg/day	0	PM2_1_0	300	26-Aug-03	28-Aug-03
100008	EX	13	Study_M 400 mg/day	Study_M 400 mg/day	100	AM1_1_100	100	29-Aug-03	29-Aug-03
100008	EX	14	Study_M 400 mg/day	Study_M 400 mg/day	0	AM2_0_100	100	29-Aug-03	29-Aug-03
100008	EX	15	Study_M 400 mg/day	Study_M 400 mg/day	0	PM1_0_100	100	29-Aug-03	29-Aug-03
100008	EX	16	Study_M 400 mg/day	Study_M 400 mg/day	0	PM2_0_100	100	29-Aug-03	29-Aug-03

Table 2. ADEX Dataset

Domain Abbreviation	Unique Subject Identifier	Sequence Number	Name of Treatment	Dose	Dose Description	Dose Units	Dose Form	Dosing Frequency per Interval	Route of Administration	Start Date/Time of Treatment	End Date/Time of Treatment
EX	100008	1	Study_M	.	AM1_1_100	mg	CAPSULE	ONCE	ORAL	2003-08-22	2003-08-23
EX	100008	2	Study_M	.	AM2_1_0	mg	CAPSULE	ONCE	ORAL	2003-08-22	2003-08-23
EX	100008	3	Study_M	.	PM1_1_0	mg	CAPSULE	ONCE	ORAL	2003-08-22	2003-08-23
EX	100008	4	Study_M	.	PM2_1_0	mg	CAPSULE	ONCE	ORAL	2003-08-22	2003-08-23
EX	100008	5	Study_M	.	AM1_1_100	mg	CAPSULE	ONCE	ORAL	2003-08-24	2003-08-25
EX	100008	6	Study_M	.	AM2_1_0	mg	CAPSULE	ONCE	ORAL	2003-08-24	2003-08-25
EX	100008	7	Study_M	.	PM1_1_100	mg	CAPSULE	ONCE	ORAL	2003-08-24	2003-08-25
EX	100008	8	Study_M	.	PM2_1_0	mg	CAPSULE	ONCE	ORAL	2003-08-24	2003-08-25
EX	100008	9	Study_M	.	AM1_1_100	mg	CAPSULE	ONCE	ORAL	2003-08-26	2003-08-28
EX	100008	10	Study_M	.	AM2_1_100	mg	CAPSULE	ONCE	ORAL	2003-08-26	2003-08-28
EX	100008	11	Study_M	.	PM1_1_100	mg	CAPSULE	ONCE	ORAL	2003-08-26	2003-08-28
EX	100008	12	Study_M	.	PM2_1_0	mg	CAPSULE	ONCE	ORAL	2003-08-26	2003-08-28
EX	100008	13	Study_M	.	AM1_1_100	mg	CAPSULE	ONCE	ORAL	2003-08-29	2003-08-29
EX	100008	14	Study_M	.	AM2_0_100	mg	CAPSULE	ONCE	ORAL	2003-08-29	2003-08-29
EX	100008	15	Study_M	.	PM1_0_100	mg	CAPSULE	ONCE	ORAL	2003-08-29	2003-08-29
EX	100008	16	Study_M	.	PM2_0_100	mg	CAPSULE	ONCE	ORAL	2003-08-29	2003-08-29

Table 3. EXDOSTXT Mapping Design in EX Domain

6. DIFFICULTY IN UNDERSTANDING THE SOURCE OR DERIVATION METHODS FOR IMPUTED OR DERIVED VARIABLES IN INTEGRATED/POOLED DATA, SUPPLEMENTAL QUALIFIERS, AND RELATED RECORDS

No such the issue happened to the case study.

7. LIMITED TRACEABILITY KEPT FROM NAÏVE DATA DEFICIENCY IN LEGACY STUDY

1) Non-supported Terminology

In one sub-study, all questionnaires in CDS-R Questionnaire form were not supported by QSTEST non-extensible code list in SDTM Controlled Terminology. After discussed with sponsor, we still kept them as user-defined code list value in QSTEST variable and explained the issue in reviewer's guide. Table 4 presents part of QSTEST code lists in Define.xml.

Clinical Global Impression Test Name [CL.QSTEST]

Permitted Value (Code)	Display Value (Decode)
CGI01-Severity of Illness [*C100955*]	CGI01-Severity of Illness
CGI01-Global Improvement [*C100956*]	CGI01-Global Improvement
CDS-R-01 I Always Feel Energetic	CDS-R-01 I Always Feel Energetic
CDS-R-02 I Am Losing Weight	CDS-R-02 I Am Losing Weight
CDS-R-03 I Have Dropped Many Activities	CDS-R-03 I Have Dropped Many Activities
CDS-R-04 Since Illness Lost Sex Interest	CDS-R-04 Since Illness Lost Sex Interest
CDS-R-05 Worried About Body Functioning	CDS-R-05 Worried About Body Functioning
CDS-R-06 Obviously Upset and Commotion	CDS-R-06 Obviously Upset and Commotion
CDS-R-07 Can Do Things On Work	CDS-R-07 Can Do Things On Work

Table 4. Part of User-defined Codelists in Non-extensible Codelist QSTEST

2) Subjects Only Existed in External Data

In one sub-study, there were 9 subjects only existed in external laboratory test result files, but not existed in all other domains, including IE domain. After discussed with study statistician and sponsor, the 9 subject records still added in DM domain and reported the issue in reviewer's guide. See Table 7 for the special tabulation.

Domain Abbreviation	Unique Subject Identifier	Subject Identifier for the Study	Subject Reference Start Date/Time	Subject Reference End Date/Time	Date/Time of First Study Treatment	Date/Time of Last Study Treatment	Date/Time of Informed Consent	Date/Time of End of Participation
DM	0005	0005						
DM	0009	0009						
DM	0010	0010						
DM	0012	0012						
DM	0016	0016						
DM	0024	0024						
DM	0027	0027						
DM	0030	0030						
DM	0044	0044						
DM	0001	0001	1999-03-22T11:30:00	1999-04-06	1999-03-22T11:30:00	1999-04-06	1999-03-15	1999-04-06
DM	0006	0006	1999-04-08T08:00:00	1999-04-26	1999-04-08T08:00:00	1999-04-26	1999-04-01	1999-04-26
DM	0003	0003	1999-03-30T12:00:00	1999-05-07	1999-03-30T12:00:00	1999-05-07	1999-03-22	1999-04-27

Table 7. The 9 Empty Subjects in DM Domain

CONCLUSION

Based on the LDCP implementation in the case study, we successfully delivered SDTM and ADaM submission packages to the client for their ISS/ISE analysis. As those examples shown, the limitations of keeping traceability are from many aspects. Other than that, we realized another challenge might come from primitive data deficiency that needs the study team to take more time to address.

To perform legacy data conversion in a more efficient way should depend on the capability of the study team and tool. The highly experienced study team is able to make difficult decision on outstanding issues, and the comprehensive issue tracking system is helpful to keep what issue happened and what decision

is made to resolve the issue; then share with other study teams as lesson learned. That will make legacy study conversion more complete and also keep traceability of every data element as originally collected.

REFERENCES

[1] U.S. Food & Drug Administration. “FDA Data Standards Catalog”. Version 4.10 (10-24-2017). Available at https://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM340684.xlsx

[2] U.S. Department of Health and Human Services, Food and Drug Administration. “Study Data Technical Conformance Guide: Technical Specifications Document”. Version 4.0, October, 2017. Available at http://www.fda.gov/downloads/ForIndustry/DataStandards/StudyDataStandards/UCM384744.pdf

[3] Dave Evans, Stephen Castellano. “CDISC Standards: A Catalyst for Industry Transformation”. November, 2016. Available at https://www.accenture.com/t20161125T001213Z__w__/us-en/_acnmedia/PDF-12/Accenture-CDISC-Standards-Catalyst-Industry-Transformation-Report.pdf

ACKNOWLEDGMENTS

The authors would like to thank Xiaojun (Maggie) Ma of for her review comments that improved the completeness of this paper.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Shengfeng (Philip) Ho
Rundo International Pharmaceutical Research & Development Co.,LTD
+86-0513-85288638 ext: 2032
shengfeng.he@rundo-cro.com / shengfeng.ho1016@gmail.com (personal)
https://www.linkedin.com/in/shengfeng-h-2b101b41/

Xiaojia (Jessie) Wang
Rundo International Pharmaceutical Research & Development Co.,LTD
+86-0513-85288638 ext: 2038
xiaojia.wang@rundo-cro.com