

Looking for SDTM Migration Specialist

Angelo Tinazzi, Cytel Inc., Geneva, Switzerland

ABSTRACT

Despite CDISC became a widely accepted standard for electronic data submission with US-FDA soon no longer accepting non-CDISC submission and other regulatory agencies considering the same (e.g. the Japan PMDA initiated a pilot phase to assess CDISC submission requirements), many studies, not only the legacy ones, are still designed without having in mind the full CDISC-path (at least CDASH-SDTM from the eCRF design step).

As a consequence migrating non-SDTM studies raw datasets to SDTM-compliant datasets is still absorbing a lot of sponsor resources. This paper discusses the implications/challenges of migrating studies to SDTM with practical and detailed examples from our experience. The goal is also to demonstrate that SDTM migration could be extremely challenging (and risky) especially with studies with complicated design either from the protocol or database design point of view (e.g. oncology studies) where in many circumstances SDTM migrated datasets represent an "extreme-derived" version of the original raw datasets.

INTRODUCTION

While the decision of going with CDISC for a regulatory submission is becoming a requirement (at least for FDA), the stage when CDISC conversion take place is still at the discretion of the single organizations; this may depend on factors such as the organizational size, where bigger does not necessarily represent a positive factor, existing company-standard, standards know-how and of course the amount of available resources [1].

In many instances the question is whether it is worth to adopt CDISC given the phase of the development; for many small companies the "wait-and-see" approach might be more appealing given their lack of expertise and the resources available.

For any organization starting from scratch, implementing CDASH should be an easy decision because it requires the development of new company-standards. However, despite some modern data-management systems come with a pre-packaged CDASH default templates allowing for some standard data domains to have data in SDTM format (when data are extracted), the "at-posteriori" SDTM conversion, or the so called legacy conversion¹, is as of today the most common approach.

The design and development of SDTM domains from legacy data requires careful planning and execution in order to produce a high quality set of deliverables. The placement of this activity introduces additional challenges (and risks) that need to be addressed well in advance [2], furthermore it requires the availability of professionals such as the "SDTM Migration Specialist".

SDTM: THE STATE OF THE ART

It has been more than ten years since CDISC published the first Stud Data Tabulation Model or SDTM [3]. Since then a number of versions have been released together with a couple of pilot studies. Furthermore the FDA, as soon as sponsors have started to submit data in SDTM format, have run a number of initiatives to share what they have seen as major misinterpretation (or gap) in the STDM data part of the submission packages they received; as an outcome

¹ A legacy conversion would be any conversion that involves sources data being captured in a given format (where not CDASH and SDTM have been used) which requires a significant transformation effort to meet downstream business needs [4].

PhUSE 2014

of such initiatives additional regulatory guidance and technical documents have been released with regards to data submission with clear reference to CDISC topics, either SDTM or ADaM.

Table 1 lists the key documents, standards or guidance, we should consider during the SDTM migration process.

Document / Guidance
Most recent SDTM released standard guidance: <ul style="list-style-type: none">- CDISC Study Data Tabulation Model (SDTM) v1.4- Study Data Tabulation Model Implementation Guide (SDTMIG) v3.2- Associated Persons Implementation Guide (SDTMIG-AP) v1.0
Released Therapeutic Area Standards: <ul style="list-style-type: none">- Diabetes, Alzheimer, Asthma, Multiple Sclerosis, Pain, Parkinson Disease, Polycystic Kidney Disease, Tuberculosis and Virology- Oncology for Tumor Response domains (integrated into SDTMIG 3.1.3)
CDISC/NCI-EVE Standard Controlled Terminology
FDA Guidance and Technical Documents: <ul style="list-style-type: none">- CDER and CBER: Guidance for Industry - Providing Regulatory Submissions in Electronic Format - Standardized Study Data (Draft)- Study Data Specifications (soon replaced by Study Data Technical Conformance Guide)- CDER Common Data Standards Issues

Table 1: The SDTM Migration Specialist Bookshelf

STEPS TO FOLLOW IN THE SDTM MIGRATION PROCESS

SDTM migration could be accomplished by following a rigorous process; this process can be divided into at least 5 main steps:

- Gap analysis
- Understanding source datasets
- Modelling the Migration
- Migration
- Finalize, Validate and Document

GAP ANALYSIS

This is probably the most important step for a successful migration and it has to be completed prior to commencing any migration activity [5]. Having a proper gap analysis does not only give an idea on how complex will be the migration, but most important it gives the possibility to the migration specialist to address well in advance potential issues and, most important, if the specialist is coming from a third party that was not involved in the study development process, it gives the possibility of making an inventory on what is available and what is not. This is extremely important with wider migration such as those required for pooling (either for efficacy, ISE, or for safety, ISS), with legacy studies conducted by different organizations (CROs), with different conventions applied and sometime in different 'era'. In some circumstances it would be not a big surprise discovering that key documents, such as the most recent CRF, are not available or that key information were not coded in the original source datasets, thus making more complicated the medical coding up-versioning (required for ISS).

A Gap analysis should address the following topics and collect the following key information:

- Itemization and evaluation of files to support migration activities
 - Study documents
 - CDISC Standards
 - Company Standards / Company Implementation Guidance
- Validate sample CRF fields versus source data
- Comparison of protocol amendments/versions against CRF versions
- External data requirements e.g. central labs
- Clarifies the scope and challenges of migration activities
- Identifies differences in data collection formats

UNDERSTANDING SOURCE DATASETS

As soon as the study material is made available to the SDTM Migration Specialist, some key decisions need to be taken with regards to the data:

- Data that needs to be submitted
- Data that can be used as it is (structure-wise)

PhUSE 2014

- Data that needs to be transposed and brought to a normalized structure
- Data that needs detailed pre-processing, including terminology migration

In this phase a first 'pruning' of the available datasets into potential SDTM equivalent dataset(s) is performed (see figure 1). Digging into the data will also give the possibility of identifying additional gaps or difficulties previously not addressed. This is often the case for laboratory data where the migration specialist may be also requested to cover processes previously not covered, such as laboratory data coming from local laboratory, thus requiring additional effort (e.g. for result conversion into SI unit or the merge with local normal ranges).

SDTM	Raw Datasets			
AE	AE	AEAER		
CM	CM	CMP	PRAML	BPP
DM	DEMO	SCEL	DS	
DS	DEMO	DS	RANDOM EOT	
EX	INFA	INFB		
IE	IE (only violated criteria)			
MH	MH	CH	PSC	DIAG
VS	VS	PMBSA		
YG	NCNC	NPTN	SCT	
ZG	CG	MARA	LVEF	NLMU
ZR	PRAM	CD	RA	

Figure 1: Data domains first 'pruning' – Multiple Sources

MODELLING THE MIGRATION

Once all the required study documents and data are available and the SDTM Migration Specialist got a better confidence with source data, it is then time to start thinking how data will or should look like once they will be converted into SDTM format. One way is to map into the original CRF the SDTM domains and variables; this is the so called Annotated CRF or blank.pdf (figure 2) where the original CRF, or the original EDC page, is annotated with the SDTM domains and variables.

DM = Demographics VS = Vital Signs

XKA46143R1 : Demography (DEMO)

Demography

1. Date of Birth | Req/Unk | Req/Unk | Req (1900-1999) | **AE** | **AEAEER**

2. Race | **RACE** | Caucasian | Black | Asian | Other, please specify : A60 | **RACE0TH** in SUPPDM

3. Sex | **SEX** | Male | Female

4. Height | **VSTEST** | cm | inch | **VSORRES** | **VSORRESU** where **VSTESTCD = HEIGHT**

Subjects: Initials

Subject ID (A60den) | A60 | **[NOT SUBMITTED]**

Subject initials | A3 | **[NOT SUBMITTED]**

Figure 2: An annotated CRF page

To avoid misinterpretations among the programming team, it is suggested to provide further specifications so that data-interpretations are not left to the single programmer. Although not always provided in the mapping package provided by the vendor, at Cytel we think this is a crucial document to take track of all transformations occurred during the migration process; the annotated CRF does not give such a 'transparency' and very often complex derivations are not documented. This document may be a complement of the Annotated CRF and it gives the possibility to the sponsor, and eventually to any auditor, to check data transformations occurred during the migration.

PhUSE 2014

This additional document may contain several sections, such as the list of source datasets contributing to each SDTM domain, the selection made on the source datasets, the mapping between legacy controlled terminology and the CDISC terminology (figure 3), complex mapping requiring input from multiple source legacy datasets (figure 4).

SDTM Variable	Raw Variable	CDISC Controlled Terminology	Original Terminology
ETHNIC	ETHNIC	<ul style="list-style-type: none"> • HISPANIC OR LATINO • NOT HISPANIC OR LATINO • NOT REPORTED 	<ul style="list-style-type: none"> • HISPANIC OR LATINO • NOT HISPANIC OR LATINO • NOT APPLICABLE (NON-US SITE)
RACE	RACEx	<ul style="list-style-type: none"> • ASIAN • BLACK OR AFRICAN AMERICAN • MULTIPLE • NOT REPORTED • OTHER • WHITE 	<ul style="list-style-type: none"> • RACE2='X' • RACE3='X' • More than one RACEx specified • RACE6='X' or RACE4='X' • RACE7='X' • RACE5='X'
SEX	SEX	F M	Female Male

Figure 3: Original controlled terminology mapped to CDISC controlled terminology

--CAT	--SCAT	--DECOD	--TERM	EPOCH	Derivation Method
PROTOCOL MILESTONE		INFORMED CONSENT OBTAINED	INFORMED CONSENT OBTAINED		DM.DSSTDATF
PROTOCOL MILESTONE		RANDOMIZED	RANDOMIZED		RD.RDDATF
DISPOSITION EVENT	END OF TREATMENT	As per CT	As per CT	TREATMENT	<ul style="list-style-type: none"> • Data from ET • DSSTDTC defined as last treatment date (DM.RFENDTC)
DISPOSITION EVENT	END OF STUDY	As per CT	As per CT	FOLLOW-UP	Data from DS

Figure 4: A complex mapping: the DS (Disposition) domain

Controlled Terminology

A key element in the SDTM migration is the adoption of the CDISC/NCI Controlled Terminology (CT) and its implementation can be complicated for several reasons. To give an example let's consider the conversion of the terminology used in a trial for collecting the route of concomitant medications as a free text; the difficulty here is not just limited to the harmonization of several terms that may have used by the different Investigators, but most important the identification of the correspondent term in the standard CDISC CT. Table 2 lists some examples where the conversion to the CDISC CT could pose some doubts to the SDTM Migration Specialist; in this case, despite the additional explanation provided by CDISC in the CT document, it may be not always self-evident that the term 'Infusion' can be assigned to 'PARENTHAL', the equivalent term in the CT, thus in this case the experience of the SDTM Migration Specialist may be not enough and a medical coding specialist should be consulted [6].

Source	CDISC/NCI CT Term (Additional explanation from the CDISC/NCI CT)
INFUSION	PARENTERAL (Administration by injection, infusion, or implantation)
MOUTH/THROAT	ORAL (Administration to or by way of the mouth)
TOPICAL-EYE	OPHTHALMIC
TOPICAL PO	TOPICAL

Table 2: Converting source text to CDISC/NCI Controlled Terminology

Trial Design

The Trial Design Model (TDM) domains can be especially challenging because, in most cases, they are being created retrospectively from the protocol, and cannot be created from electronic data; these domains are meant to define the design of a clinical trial in terms of Arms and schedule (e.g. visits). This is probably the first step in the

PhUSE 2014

migration and should be defined even prior to start the migration specifications. As discussed in [7] two or more individuals may read a protocol and design the TDM differently; I invite all the readers of this paper to look at the good tutorials of Wood and Lenzen.

In some recent SDTM migration experiences one domain in particular that caused some issue was the Trial Summary (TS). TS is the domain that includes key details, called Parameters, about the trial that would be useful in clinical trials registry; for example the title, objectives, minimum age, maximum age, and indication. We found the identification of key details to be included fairly documented in the various SDTM IG versions and very often we got several issues while validating SDTM with Open CDISC, but finally based on the errors/warnings from OpenCDISC we got an understanding of which parameters was needed. See also [8].

New domains or sponsor domain

Although SDTM IG 3.2 contains more new domains than previous versions, one of the most critical decisions for the SDTM Migration Specialist is whether or not it is required to create a new domain, non-standard domain, or the so called “sponsor domain”. After having verified that neither a standard SDTM domain exists or data could be not mapped into an existing domain with appropriate category definition (--CAT or –SCAT), the following main steps should be followed in the creation of the domain:

- Use existing models as prototype or choose the appropriate general observation class
- How to determine where data belong in the SDTM. See IG v3.1.3 section 8.6
- Assign 2–letters domain code. Possible conventions:
 - Z- for findings
 - Y- for interventions
 - X- for events

Figure 5 and 6 provide some more details on the process to be followed for the creation of a new domain. The example is about anti-cancer therapies where we decided to create a separate domain and not storing this information in CM with all other ‘standard’ medications. Of note some specific oncology procedures, such as radiotherapy and surgery, may be mapped to the new PR domain available since IG 3.2.

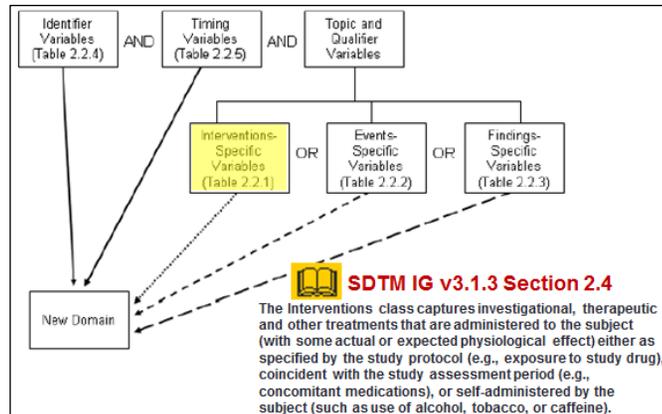


Figure 5: the process for creating a sponsor domain

IDs 2.2.4		Timing 2.2.5				
Variables which identify the study, the subject, the domain and the seq nr. of record		Variables which describe the timing of an observation (e.g. date collection)				
STUDYID	USUBJID	DOMAIN	YGSEQ	YGSPID	YGSTDTC	YGENDTC
XXXX	YYYYYY	YG	1	1	2011-01-12	2011-02-20
XXXX	YYYYYY	YG	2	1	2011-01-12	2011-02-25

Interventions Qualifiers 2.2.1		
Select and include the relevant Qualifier variables from the identified general observation		
YGTRT	YGCAT	YGSCAT
FTC PLUS SORAFENIB	SUBSEQUENT NON-PROTOCOL AML TREATMENTS	CHEMOTHERAPY
CLAG-M THERAPY	SUBSEQUENT NON-PROTOCOL AML TREATMENTS	CHEMOTHERAPY

Figure 6: applying the process for creating a new domain

PhUSE 2014

MIGRATION

Once the SDTM migration has been modeled and therefore all specifications available, a programming team can be assigned. Several strategies are available depending on the type and amount of studies to be migrated. For example if the main purpose of the SDTM migration is to support either an ISS or an ISE, it may make sense to assign programmers by data domain, so that for example all AEs are programmed by the same set of people.

FINALIZE, VALIDATE AND DOCUMENT

A number of activities have to be performed as a final step in the migration process; these activities are either for completing the mapping or to prevent later issues during the validation with OpenCDISC [9]. These activities are related but not limited:

- Creation of domains containing actual protocol events (Subject Visits-SV and Subject Element-SE) from migrated SDTM datasets
- Identify key variables determining sorting order of each domains and therefore --SEQ generation
- Remove duplicates such as repeated 'Not Done' records belonging to the same event/sample. This could be the case is the case of an entire lab assessment not performed; as shown in the example in figure 6 we have replaced all single 'Not Done' lab parameters with one record saying the entire sample at that specific time-point was 'Not Done'

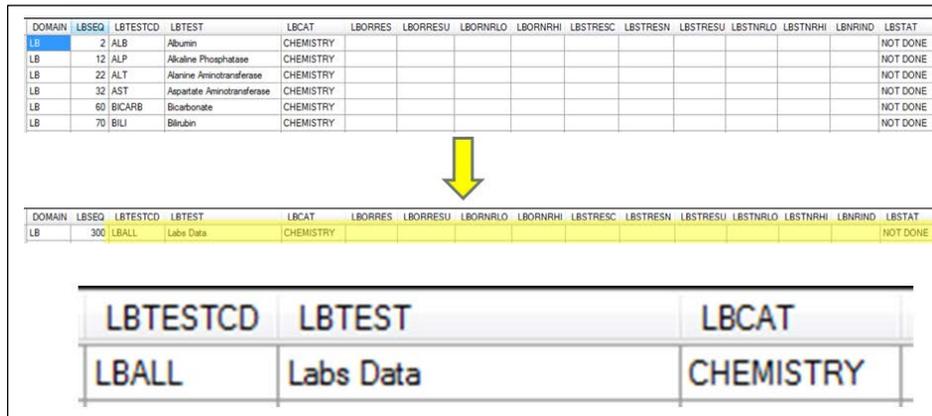


Figure 6: the process for creating a sponsor domain

- Reduce variables length: one of the issue raised by FDA CDER with the Common Issues document, was the size of datasets they received. Length/size for each variable should be not arbitrarily set to 200 and therefore it is suggested as ultimate step in the migration to reset the variables length of each variable based on the actual (final) variables length. Figure 7 below provide an example with AETERM in the AE domain

Variable length should be assigned based on actual stored data to avoid to minimize file size

AETERM

H	Y	P	E	R	K	A	L	E	M	I	A				
---	---	---	---	---	---	---	---	---	---	---	---	--	--	--	--

Length=14

H	Y	P	E	R	K	A	L	E	M	I	A
---	---	---	---	---	---	---	---	---	---	---	---

Length=12

- **Exception for**
 - All variables with identical names in different SDTM domains e.g. USUBJID, STUDYID, VISIT, EPOCH
 - --DTC variables (always \$19)

“Warning” issued by Validation Check SD1081

FDA CDER Common Data Standards Issues Document

Figure 7: An example of variable length re-size based on actual data length

PhUSE 2014

- **Split Domains:** in some circumstances the size of the datasets can increase due the number of records. In this case a split of the domains could be the solution. It is suggested to make the split based on records categories, such as –CAT; for example Laboratory category, having an LB domain for haematology and one for chemistry

For certain aspects SDTM can be only finalized at the time of db-lock. For example:

- Medical Coding to be completed
- Assign visit number to “Unplanned Visits”. This is often the case for oncology studies, for example Phase I, where the number of visits are not known at-priori
- Some Controlled Terminology can be only finalized at the time of final db-lock. This is often the case for laboratory units especially with trials using local laboratories

Validate

There are two levels of validation. The first one is related to the validation or better to the Quality Control (QC) of the programming; this kind of validation should follow the same principles as any other validation programming activities (e.g. for ADaM or for analysis outputs, applying independent double programming when required). It is also suggested to re-program some key statistical outputs using the migrated SDTM datasets then comparing the same outputs created with original legacy data.

The second level of validation is usually performed against the CDISC SDTM standard validation checks implemented in OpenCDISC; these checks make sure the mapping was done according to the criteria proper of the chosen SDTM version and some of these checks may also identify potential data issues.

Document - The Study Data Reviewer Guide

This is a crucial final step where not only unresolved, or unresolvable, issues raised by the OpenCDISC validator have to be documented and eventually explained to the reviewer, but it is also the document where the SDTM Migration Specialist can provide additional details about the mapping, such as the identification of key efficacy components (e.g. where is located the primary efficacy endpoint).

Although not yet mandatory it is recommended to use the template provided by the PhUSE/CSS initiative [<http://www.phuse.eu/CSS-deliverables.aspx>].

Figure 8 provides a couple of examples of key additional information that can be provided in the reviewer guide:

- The first one is about randomization factors placed in the ‘basket’ of DM supplemental qualifier; these are important variables to be used in the efficacy analysis and the SDTM Migration Specialist did not find a better location, but given the importance the location of the variables was highlighted in the reviewer guide.
- The second comment instead is an anticipation on how the key efficacy endpoint, the overall survival (time-to-death), should be derived under specific situations, either when the patient died or was lost to follow-up.

Key analysis data points include:

- **Randomization strata are in SUPPDM**
 - Age: QNAM= AGEGROUP
 - Disease Status: QNAM= DISSTAT
 - Geographical Region: QNAM= LOCATION

- **Primary Efficacy Endpoint:** death subjects are in DM where DTHFL=Y, with date of death DM.DMDTC and cause of death in SUPPZS where QNAM=DEATH. For subject alive, last follow-date should be retrieved from ZS; for statistical analysis purpose last date patient was seen alive should be calculated using any date available in the SDTM datasets, except for the following variables:
 - ZS.ZSSTRESC=DEATH, in this case the date (ZSDTC) may be after date of death as this date represents the date when the investigator was informed of the death. For the same reason is not recommended to use date in SV dataset
 - DM.DMDTC

Figure 8: Some key information that can be provided in the reviewer guide

PhUSE 2014

CASE 1: GAP ANALYSIS

In 2014 we have been approached by several potential customers requiring our consultancy for SDTM migration, and more in general for submission, asking to make an assessment of their studies materials (prior to sign any contract agreement) in order to identify possible gaps and assess the resources required for the submission support.

One customer in particular, a small biotech with limited operational experience and internal programming/biostatistics resources, wanted to complete an ISS and an ISE submission within three months. In order to assess the feasibility of such timelines, the customer agreed to have Cytel assessing the status of the projects, by reviewing the status of the studies and identifying key characteristics, assessing the available study materials and most important assess the data migration status of each single study and eventually quantify the gap where the migration was not yet performed.

We soon discovered there was a lack of understanding of some key concepts such as traceability; for example the client was not even aware of differences between raw datasets to migrate to SDTM and Analysis datasets that may have integrated some raw datasets.

Figure 9 below, and figure 10 in appendix, provides more details about information that have been collected during the gap analysis. The information collected will be useful for assessing the current status of the SDTM migration and therefore decide whether or not a version alignment is necessary, to make a first assessment of the raw datasets for trials where migration to SDTM was not already conducted, assess the status of the medical coding and the availability and complexity of external data.

Protocol	Phase	Ongoing/ Closed?	Study Subjects	Country	Raw Data	SDTM	Analysis Data	protocol	SAP	aCRF	Blank CRF	CSR
TRIAL 01	Ib	Closed	CF with Pa	Netherlands	present	To migrate	Non-ADaM	v2.0.1 / 2006	v1.1 / 2007	v1.7.4 / 2007	NA	v1.1 / 201301
TRIAL 02	Ib/Ila	Closed	CF with Pa	Hungary, Serbia	Waiting- will get from another vendor	To migrate	Non-ADaM	v1.1 / 2007	v2.0 / 2010	v1.1 / 2007	NA	v1.1 / 201404
TRIAL 03	Ib/Ila	Closed	CF with Pa	Belgium	present	To migrate	Non-ADaM	v1.4 / 2008	v1.2 / 2010	v2.2 / 2008	NA	v1.1 / 201304
TRIAL 04	III	Ongoing	CF with Pa	North America, Europe, Australia, New Zealand	present	v3.1.3	Available	v1.4 / 2009	v.2.1 / 2011	v1.5 / 2010	Available	Not Applicable

Figure 9: Gap Analysis – Documents Inventory

CASE 2: AN ONCOLOGY STUDY WITH A COMPLEX STUDY AND CRF DESIGN

This was an oncology phase III study and it is a perfect example to demonstrate how complex could be a migration in presence of a complex protocol and CRF design. Moreover the high number of recruited patients (about 700 patients) increased the number of scenario requiring “ad-hoc” programming activities. Last but not least the number of external files to incorporate, either excel file or SAS datasets, such as normal ranges and conversion to SI lab units, the medical coding and the data requiring expert and independent assessment (efficacy data),

Most of the examples provided in this paper and the one that will be provided during the presentation are coming from this study.

CONCLUSION

Legacy SDTM conversion is a complicated task that requires careful forethought and planning from the very beginning. As an initial step it is important to assess the entire set of study documents, including the legacy data, in order to identify any possible gap and set a clear ground for starting the migration.

PhUSE 2014

While someone may still think SDTM migration can be easily achieved by simply renaming and adapting data, there are many aspects that require interpretation, clear knowledge of the data and the clinical protocol. Very often decisions have to be taken with regard to therapeutic area specificity; this is proven by the number of recurrent questions on the different CDSIC channels and forums (for example with oncology data) [10]. The relatively new concept of developing CDISC therapeutic area Guidance may cover a lot of the gaps we have seen in our experience.

More and more we see the figure of the SDTM migration specialist a “must” for all organizations, either sponsors or CROs wanting to either submit data into SDTM format or why not reshape the entire data-management standards more towards the CDISC standards.

REFERENCES

- [1] H Winsor, Good versus Better SDTM – Why “Good Enough” May no Longer be Good enough when it some to SDTM, PharmaSUG 2014
- [2] DC IZard, SDTM Domains development on the Critical Path - Strategies for Success, PharmaSUG 2012
- [3] SR Lee, Experiences in Leading a Company-wise first CDISC filling from a Programming Perspective, PharmaSUG 2012
- [4] F Wood, Creating SDTM Datasets from Legacy Data, PharmaSUG 2011
- [5] I Dark, J Negrelli, A Sathe, New CDISC migration strategies and tools: determining the best approach and employing automated data mapping, PhUSE 2010
- [6] MR Martinez, A how-To Guide for Extending Controlled Terminology Using SAS Clinical Data Integration. PhUSE 2013
- [7] F Wood, M Lenzen, Trials and Tribulations of SDTM Trial Design. PharmaSUG 2011
- [8] N Tambascia, A Esiberg. A “How-to Guide” for Trial Summary 3.1.3. PhUSE SDE Frankfurt 2014
- [9] F Dilorio, OpenCDISC: Beyond Point and Click, PharmaSUG 2013
- [10] A Tinazzi, Therapeutic Area Standards: Reflections on Oncology standards and what is needed in the future, CDISC Interchange Europe 2014
- [11] P Nacci, Legacy Data Conversion: A Journey of Discovery, PhUSE 2012
- [12] C Holland, J Shostak, Implementing CDISC Using SAS, SAS 2012

ACKNOWLEDGMENTS

I would like to thank Dean Shults of the Cytel Geneva offices for their in-depth review and for providing valuable comments and suggestions.

CONTACT INFORMATION

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

Angelo Tinazzi

Cytel Inc.

Route de Prè-Bois 20

1215 Geneva – Switzerland

+41 765359946

Email: angelo.tinazzi@cytel.com

Web: www.cytel.com

Brand and product names are trademarks of their respective companies

PhUSE 2014

Item	Studies			
	TRIAL 01	TRIAL 02	TRIAL 03	TRIAL 04
ISS/ISE	ISS	ISS	ISS	ISS/ISE
Availability of the Protocol	v2.0.1 / 2006	v1.1 / 2007	Available from CSR v1.1 / 2007	v1.4 / 2009
Availability of the Annotated CRF	Available	Available	Available	Yes From the aCRF the raw datasets seems very closed to SDTM, a kind of SDTM+ (e.g. sponsor domains do not start with X,Y,Z)
Availability of the Raw Data (add the list of files)	Yes But to confirm if this is the final source since the datasets seems a copy/derivation of another db extract	Yes Only PK plus some labs, demo and drug admin	Yes	No Derived datasets are a mix between raw data and derivations
If raw data are available, is it complete? Compliant with the annotated CRF?	NA	Only for PK	No	No. See comment above.
Availability of SDTM	No	To come from another CRO	No	No
Availability of the ADaM/Derived Please add the list of files	Available (16 Analysis datasets), this seems however copy-re-arrangement of other raw data: AE, AECODE, COMMENT CONMEDS, HEMACHEM MEDHIST, MICROBIO PATIENT, PE, PKSERUM PKSPUTUM, PKURINE PREG, SPIRO SPUTUM, VITALS	No	Yes. Non ADaM	ADHOCLABS AE AUCTC AUDIOLOGY CFQR CFQRRAW CHESTXRAY COMPLI CONMED FEV FEVMAX LAB MICRO O2SAT PATDISP PATIENT PE PKSERUM RANTIB SCREENFAIL SPCHANGE SPCHGEXPLORE STOPRULE VS VSCHANGE WEIGHTCHG Derived datasets specifications available RANDASSI described in the documentation is not available; however most of the datasets have the assigned random variable available

PhUSE 2014

	Studies			
Item	TRIAL 01	TRIAL 02	TRIAL 03	TRIAL 04
Status of AE coding	Coded. Coding variables available: PT_AECODE PT_NAME PT_SOC_CODE SOC_NAME MedDRA v 9.0	Un-coded	Coded. Coding variables available: PT_AECODE PT_NAME PT_SOC_CODE SOC_NAME MedDRA v 12.1	PT/SOC Text available. No LLT. No MedDRA coding variables available (e.g. LLT code) Unknown MedDRA version used for coding
Status of ConMed coding	Not coded	Not coded	Not coded	Class/PT available. Unknown level of coding used for class. WHO-DD version to be clarified
Status of labs data: Local labs, central labs, SI units ?	SI results available	Chemistry only available for now but SI unit conversions seems available for all parameters	SI results available	Central labs. SI unit available, although they make use of US conventional unit in some cases (e.g. ALAT ukat/l instead of U/L)
List of lab Parameter	Chemistry: Chemistry: Albumin, ALAT, ALKPhos, ASAT, Bilirubin, Calcium, Chloride, Cholesterol, GGT, Glucose, HDL, Uric Acid, Urea, Potassium, Sodium, Phosphate, Triglycerides Hematology: WBC, Neutrophils, Basophils, Eosinophils, Monocytes, Platelets, Hemoglobin, RBC, Hematocrit	Chemistry: Chemistry: Albumin, ALAT, ALKPhos, ASAT, Bilirubin, (total), Calcium, Chloride, Cholesterol, Creatinine (clearance and serum), Potassium, LDL, Sodium, Phosphate, Total Protein, Triglycerides Hematology: WBC, Neutrophils, Basophils, Eosinophils, Monocytes, Platelets, Hemoglobin, RBC, Hematocrit Coagulation: MCH, MCHC, MCV	Chemistry: Chemistry: Albumin, ALAT, ALKPhos, ASAT, Bilirubin, (total / direct), Calcium, Chloride, Cholesterol, Creatinine (clearance and serum), GGT, Glucose, HDL, Uric Acid, Urea, Potassium, Sodium, Phosphate, Total Protein, Triglycerides Hematology: WBC, Neutrophils, Basophils, Eosinophils, Monocytes, Platelets, Hemoglobin, Hematocrit Coagulation: MCH, MCHC, MCV	Chemistry: Chemistry: Albumin, ALAT, ALKPhos, ASAT, Bilirubin, (total / direct), Calcium, Chloride, Cholesterol, Creatinine (clearance and serum), GGT, Glucose, HDL, Uric Acid, Urea, Potassium, LDL, Sodium, Phosphate, Total Protein, Triglycerides Hematology: WBC, Neutrophils, Basophils, Eosinophils, Monocytes, Platelets, Hemoglobin, RBC, Hematocrit Coagulation: MCH, MCHC, MCV
Status of the PK/PD data	Available report and derived datasets	Report available	Report available	PK and Bioanalytical Reports available. Derived datasets

PhUSE 2014

Item	Studies			
	TRIAL 01	TRIAL 02	TRIAL 03	TRIAL 04
				available
Availability of Protocol deviation data . How it is incorporated?	Not available. However, based on the analysis dataset specifications, the PP population has been derived based on the raw data.	Sponsor to provide	Sponsor to provide	Not applicable, final study assessment not yet occurred
Availability of ADs specifications	Available: DataDictionary.doc (..\GAP Analysis\Analysis Datasets\TRIAL01\Datasets and Documentation)	Sponsor to provide	Sponsor to provide	Available: DataDictionary.doc (..\GAP Analysis\Analysis Datasets\TRIAL04\Datasets and Documentation)
SAP	v1.1 / 2007	v1.1 / 2007	v1.1 / 2007	v1.1 / 2007
Clinical Study Report	Available	Available	Available	Yes including section 14 and 16.1 / 16.2
Tables	Not available. It seems that only listings were done.	Available	Available	Available
Listings	Section 16.2 available (... \GAP Analysis\TRIAL01\101 TLFs)	Available	Available	Available
Figures	Not available	Available	Not available	Not available

Figure 10: Detailed Gap Analysis