

CDISC ITA UN 2020

Training
Milano, September 9-15 2020

define-xml and submission ready documents Q&A

Silvia Faini, LivaNova
Milano – Italy
silvia.faini@livanova.com

Angelo Tinazzi, Cytel Inc.
Geneva – Switzerland
angelo.tinazzi@cytel.com

? Cosa viene incluso nelle cartelle 'split'

Qualora un dataset, di solito SDTM, supera le dimensioni massime «permesse» da FDA (ad oggi 5GB), la richiesta è di fornire dataset di dividere il dataset in piu part. Per esempio per LB, se >5GB, potremmo dividere il dataset in base alla categoria es. **LBHM** per dati ematologici, **LBCH** per chemistry, **LBUR** per urinalysis, etc. Questi 'split' dataset vanno forniti nella cartella **/split**, mentre il dataset completo, LB in questo caso, va comunque fornito nella cartella principale.

Allo stesso modo, per QS, se si decide a prescindere dalle dimensioni, di avere un dataset per questionario, vale la stessa regola, QS nella cartella principale e per esempio QSSF, QSCGI, etc. In /split

? ADaM programs sono necessari (mandatory) nel modulo m5 o solo raccomandati

? Mi pare di capire che non è obbligatorio fornire anche programma SAS, corretto

Almeno i programmi di ADaM sono richiesti da FDA e suggerito fornire anche i programmi che supportano almeno il primary efficacy endpoint

? Variable Order assegnato automaticamente usando P21 non rispecchia sempre richiesta SDTM-IG. va dunque cambiato a mano

Il campo Order è un campo/metadata presente nel template P21. Nel generare define-xml P21 utilizza l'Order specificato. ATTENZIONE: Pinnacle mantiene l'ORDER specificato nel file excel, ma non riordina le variabili in base al campo ORDER. In define-xml le variabili verranno presentate secondo l'ordine con cui le variabili compaiono nel file excel

? Esempi di Good define-xml

- Alcuni suggerimenti nella seconda parte
- PhUSE Best Practices for Documenting Dataset Metadata - Define-XML versus Reviewers Guide
- PhUSE Define-XML 2.0 Completion Guidelines V1.0

? Esempi di Good define-xml (cont)

“PhUSE Define-XML 2.0 Completion Guidelines V1.0” vs “CDISC “define.xml 2.0 specifications” es. descrizione attributo “Origin”

→ PhUSE Define-XML 2.0 Completion Guidelines V1.0

Origin Type	Metadata Item	Usage Notes	Examples
Assigned	Origin Type	<p>Origin Type = "Assigned" for variables that are filled by individual judgment (by an evaluator other than the subject or investigator), rather than collected as part of the CRF, eDT or derived based on an algorithm. This category includes variables that are populated from an external dictionary.</p> <p>No other metadata items are required for Assigned origins.</p> <p>See section 2.4.2.1 Variable-level Metadata, OriginType</p>	<p>Variables related to medical dictionaries: CMDECOD, AEDECOD</p> <p>Numeric counterparts of primary variable in character format: SEXN, RACEN, AVISITN</p> <p>Character counterparts of primary variable in numeric format: APERIODC</p> <p>BDS parameter related variables: PARAM, PARAMCD, PARAMN.</p>
Derived	Origin Type	<p>Origin Type = "Derived" for variables that are calculated by an algorithm or reproducible rule that is dependent upon other data values. Derived variables must reference</p>	<p>RFSTDTC: RFSTDTC = first date/time of study drug, for safety subject.</p> <p>CMENDY: End date minus DM.RFSTDTC. Add 1 if >=</p>

? Esempi di Good define-xml (cont)

Per ADaM

→ ADAM Traceability Example (draft)



Version 1.0 (Draft)
Prepared by the
CDISC Analysis Data Model Team

Notes to Readers
This Analysis model uses the principles, structures and standards described in the CDISC Analysis Data Model: Version 2.1 and Implementation Guide v1.1 documents.

Revision History

Date	Version
2014-03-25	1.0 Draft

© 2014 Clinical Data Interchange Standards Consortium, Inc. All rights reserved.

Contents

1. Introduction
2. Purpose
3. The Need for Traceability in ADaM
4. Points to Consider When Interpreting this Document
5. Caveats and Disclaimers
6. Conventions Used in this Document
7. Examples
8. General ADaM Traceability
9. Analysis Need 1
10. Data Flow 1
11. Traceability Metadata 1
12. Input and Analysis Data 1
13. Other Uses 1
14. General OCCG Traceability
15. Analysis Need 2
16. Data Flow 2
17. Traceability Metadata 2
18. Input and Analysis Data 2
19. Other Uses 2
20. Traceability with Parameters from Multiple Input Datasets
21. Analysis Need 4
22. Data Flow 4
23. Traceability Needs 4
24. Input and Analysis Data 4
25. Other Uses 4
26. Traceability when multiple input datasets are stacked to create OCCG
27. Analysis Need 5
28. Data Flow 5
29. Traceability Needs/Metadata 5
30. Input and Analysis Data 5
31. Traceability When Multiple Datasets Are Merged
32. Analysis Need 6
33. Data Flow 6
34. Metadata 6
35. Input and Analysis Data 6
36. Traceability when Creating Rows
37. Traceability When Adding a Row to a Basic Data Structure Dataset
38. Analysis Need 7
39. Data Flow 7
40. Traceability Metadata
41. Input and Analysis Data 7
42. Other Uses 7
43. Using an Intermediate Dataset for BDS Traceability

? Qual'è il flusso che porta alla creazione del define

Dipende.....

- **Opzione 1: Creare il define (e documentazione annessa) «retrospettivamente»**

1. Crei i SAS datasets, SDTM o ADaM ... SAS7BDAT → XPT
2. Dai dataset SAS generi in automatico con P21 le specifiche in excel
3. Completati le specifiche es. Codelist, Descrizioni derivazioni, Ulteriori attributi non catturabili dai dataset SAS es. 'Mandatory', Origine, etc.
4. Generi define-xml da specifiche excel
5. Validi define e datasets+define
6. Correggi specifiche define e/o datasets
7. Torni al punto 4-6 fino a quando hai risolto tutte le discrepanze

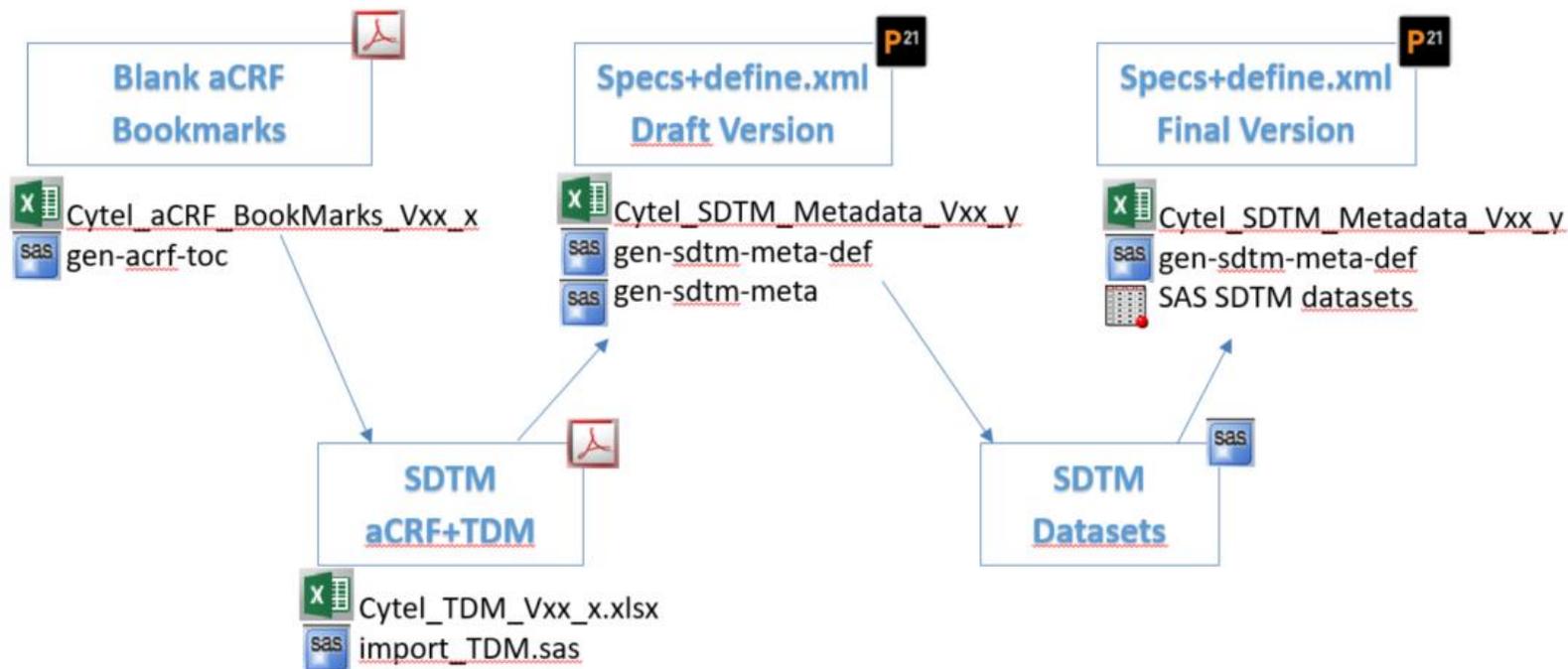
? Qual'è il flusso che porta alla creazione del define (cont)

Dipende.....

- **Opzione 2: Creare il define** (e documentazione annessa) **«prospettivamente»**
 - Per Esempio in Cytel

? Qual'è il flusso che porta alla creazione del define (cont)

A Cytel per esempio per SDTM (molto simile per ADaM)



? Nella definizione degli standard viene definito anche lo stato del documento.
<def:StandardOID="STD.IG.SDTM32" Name="SDTMIG" Type="IG" Version="3.2" Status="Final"> questo significa che è possibile utilizzare anche standard con status non «Final»?

NB. L'attributo Status è disponibile con define-xml v2.1

No / Dipende...

- **No se lo standard non è ufficiale** es. No draft, no versioni/standard rilasciate da CDISC per public review
- **No se una versione final non è ancora accettata da FDA** (o altri enti) es. SDTM IG 3.3 e ADaM IG 1.2 non sono ancora accettate

? Nella definizione degli standard viene definito anche lo stato del documento.

`<def:StandardOID="STD.IG.SDTM32" Name="SDTMIG" Type="IG" Version="3.2"`

`Status="Final">` questo significa che è possibile utilizzare anche standard con status non «Final»? (cont)

No / Dipende...

- **Dipende** - Domini SDTM non disponibili nella IG che stiamo utilizzando, es. 3.2, ma il cui «nome» è stato «riservato» nella controlled terminology, possono essere usati anche se non descritti nella IG di riferimento
- **Dipende** - Suggerimenti di implementazione disponibili in versioni draft o disponibili in nuove version, possono essere utilizzati se non contraddicono quando detto nella versione usata (Cautela!!)

? Nel caso di sviluppo di variabili ADaM da variabili RAW, non mappate in SDTM, come conviene documentarle? Nel caso sia possibile, si possono allegare i dataset “source” nel folder MISC (sempre in formato xpt), ma se per esempio queste variabili RAW non potessero essere né riportate in SDTM e nemmeno allegate alla submission dal dataset originale, avete qualche suggerimento su come approcciarsi?

Per esempio se volessimo utilizzare per un analysis le variabili generate e usate dal sistema di randomizzazione (quindi variabili non collezionate in CRF e non riportate in SDTM, ma disponibili in alcuni particolari RAW datasets).

Se avete fatto SDTM il consiglio è sempre quello di mappare queste variabili in SDTM; in SDTM non vanno solo le informazioni raccolte in CRF, ma l'SDTM deve includere anche informazioni che arrivano da central labs, IVRS (random e assegnamento farmaco se fatto con IVRS), etc. Nel caso della random list con info di stratificazione, il braccio sicuramente viene mappato in DM, mentre le variabili utilizzate per stratificare la random se non hanno una collocazione ufficiale in SDTM, di solito vengono messe in SUPPDM.

? Ho una lista di AESI (Adverse Events of Special Interest) che sono quelli di interesse per l'analisi. Ho creato un AXAE dataset che la contiene? È corretto mappare così queste informazioni? Nel campo origin del define per ADAE utilizzo «predecessor» o «assigned»?

Prima di tutto l'origine del dato è sicuramente «assigned».

La soluzione di mappare questa lista di AESI in AXAE è una delle opzioni.

Queste informazioni esterne possono essere considerate dei lookup dataset che non contengono dati legati direttamente ai pazienti. In questo caso non siamo «costretti» a trovare una collocazione in SDTM/ADaM, poichè le informazioni diventano utili quando creiamo un ADaM dataset, quindi devono essere incluse nella sottomissione.

È possibile quindi includere questo dataset (sempre in formato xpt) nel folder MISC (From «FDA Data Technical Conformance Guide» MISC definition: place miscellaneous datasets that don't qualify as analysis, profile, or tabulation datasets in this subfolder).

Ad ogni modo la cosa fondamentale resta documentare in define/ADRG in modo chiaro.

? Per le ISE e le ISS, nel folder di analysis dell'm5, conviene approcciarsi con la creazione di ADaM adhoc e nominarli con la nomenclatura standard (e.g. ADSL) oppure usare l'approccio del paper sotto e quindi chiamare i nuovi dataset I-ADaM (e.g. IADSL)?
<https://www.pharmasug.org/proceedings/2018/DS/PharmaSUG-2018-DS03.pdf>

ieri nelle slide (#153) si riportavano i nomi "usuali" per ADaM, quindi mi chiedevo se per vostra esperienza o conoscenza sapete se FDA o PMDA hanno delle preferenze a riguardo

"Purtroppo" quella proposta non è passata. Ufficialmente non è passata perchè necessita di piu' revisione interna a CDISC (ma in realtà sembrerebbe non aver riscontrato i favori di FDA). Comunque no, a prescindere dalla proposta dell'articolo che hai riferenziato, i datasets dovrebbero essere comunque nominati secondo gli standard ADAM, ovvero cominciare con 'AD'. iADSL, iBDS, iOCCDS, sarebbero dovute essere delle nuove classi ADAM per permettere di avere, nel caso di integrazioni di dati da diversi studi, in ADSL piu' di un record per soggetto, in BDS and OCCDS piu' di un record per soggetto per coorte di analisi. Purtroppo il tema 'pooling' nel mondo CDISC resta un argomento non ancora ben coperto dagli standard e ognuno alla fine fa un po' come vuole.

? Come CRO, stiamo muovendo ora i primi passi nell'ottica di poter fornire il servizio "define xml creation e submission package", perché ci siamo resi conto che i nostri clienti abituali, anziché appoggiarsi ad altre CRO per questi aspetti, preferirebbero che il servizio fosse interamente fornito da noi a corredo della creazione di SDTM, ADAM e TLFs. In questo framework, si inserisce la richiesta di una CRO partner che sarebbe incaricata della creazione del Define XML ma ci chiede di fornire delle specifiche a supporto di SDTM e ADAM (creati da noi) per aiutarli nella creazione del define xml.

Cosa ci consigliereste di fornire, considerando che sarebbe la prima volta che ci avviciniamo alla questione? i metadati? oppure le csdrg e adrg?

Di solito, come voi con l'altra CRO, riceviamo i datasets e da lì ci costruiamo noi i metadata con dei nostri programmi.

Non avendo fatto noi i datasets, chiediamo di fornirci le specifiche delle variabili derivate e anche i programmi, nel caso la documentazione delle derivazioni non fosse sufficiente.

Quindi direi almeno le specifiche delle derivazioni e poi per SDTM l'aCRF se l'avete fatta.

Poi ovviamente dipende da che strumenti la CRO utilizza per generare il define-xml e dagli agreement che avete preso con loro.

? la aCRF la creiamo già quindi quella sicuramente possiamo fornirla. Mentre per le specifiche delle derivazioni in quale formato possono essere fornite?

Le specifiche delle derivazioni possono essere fornite in un qualsiasi formato che la CRO è poi in grado di trattare (quindi non word o PDF), ovvero unire ai metadata che saranno in grado di derivare automaticamente dai dati che forniti.

Un semplice foglio excel con una struttura concordata con la CRO potrebbe essere sufficiente. Per esempio se devi fornire le specifiche delle derivazioni:

Dataset	Variable	Derivazione
---------	----------	-------------

ADSL	SAFFL	Defined as...
------	-------	---------------

Ad ogni modo avere il processo completamente interno con la produzione di tutta la documentazione in parallelo alla produzione di SDTM/ADaM/TFL fornisce garanzie maggiori sulla corretta tracciabilità dei dati.

? “Key variables” nello sheet “Datasets”: nel caso in cui le “key variables” non identificavano univocamente i records (i.e. STUDYID,USUBJID,AEDECOD,AESTDTC,AETERM per il dominio AE), è possibile aggiungere anche chiavi “teoriche” (prese magari dalla SDTM-IG e.g. AESEQ) o è necessario aggiungere variabili fino ad ottenere un record univoco?

Per definizione le key variables devono permettere di individuare univocamente un record.

In generale non può essere aggiunto il Sequence Number (e.g. AESEQ), poichè solitamente la lista della key variables è quella che aiuta la creazione di questa chiave surrogata.

In casi limite può essere aggiunta alla lista delle key variable lo Sponsor ID (e.g. AESPID), anche se la scelta è da considerarsi border-line.