

KSFE-10

Migration von klinischen Studien nach CDISC-SDTM

Claudia Meurer
Accovion GmbH
Helfmann-Park 2
65750 Eschborn
Claudia.meurer@accovion.com

Zusammenfassung

Der klinische Datenstandard CDISC-SDTM wird von der amerikanischen Zulassungsbehörde FDA empfohlen. Die Struktur des CDISC-SDTM-Modells wird vorgestellt. Um klinische Daten in diese CDISC-Struktur zu überführen, wurde bei der Accovion GmbH ein SASTM-Macro entwickelt. Dieses liest Metadaten aus ExcelTM-sheets, in denen alle zu erstellenden Datensätze mit den entsprechenden Variablen definiert sind, und erstellt daraus CDISC-SDTM-Datensätze.

Schlüsselworte: CDISC, SDTM, ADaM, ExcelTM, SASTM-Macro.

1 CDISC allgemein

CDISC (Clinical Data Interchange Standards Consortium) hat sich zum Ziel gesetzt, standardisierte Datenmodelle für klinische Daten zu entwickeln. Das Konsortium besteht aus Teilnehmern verschiedener Arbeitsgebiete (u.a. Pharmaindustrie, Biotechfirmen, Gesundheitswesen, Clinical Research Organisation). Es arbeitet eng mit der FDA (Food and Drug Administration) zusammen, die die Zulassung neuer Medikamente auf dem US-amerikanischen Markt regelt.

Verschiedene Arbeitsgruppen innerhalb CDISC beschäftigen sich mit der Definition unterschiedlicher Datenmodelle. Diese Standardisierung soll dazu beitragen, die Einreichung klinischer Studien bei Behörden zu vereinfachen

2 CDISC - SDS

Grundlage für die Standardisierung klinischer Daten ist SDS (Submission Data Standards), der von CDISC entwickelt wurde. Organisation, Struktur und Formatierung der klinischen Daten für die SDTM (Study Data Tabulation Model)- und AdaM (Analysis Data Model)- Datensätze sind hierdurch genau festgelegt.

Die klinischen Daten sind inhaltlich in den sogenannten „Domains“ zusammengefasst. Eine Domain entspricht einem SASTM-Datensatz bzw. einer Datenbank-Tabelle. So werden z.B. demografische Daten einer klinischen Studie in der Domain DM, Nebenwirkungen in der Domain AE zusammengefasst.

3.1 Domain-Kategorien

Es gibt folgende Domain-Kategorien:

- „Interventions“: sie beinhalten alle Medikamente, Therapien, Operationen eines Patienten während der Studie, z.B: Studienmedikation (EX), Begleitmedikationen (CM)
- „Events“: hier werden alle Ereignisse, die vor oder während der klinischen Studie auftreten, gespeichert, z.B: Nebenwirkungen (AE), Patientendisposition (DS), Operationen und Vorerkrankungen (MH)
- „Findings“: dies sind die Beobachtungen, die an geplanten Untersuchungsterminen gemacht werden, z.B: Labor (LB), EKG (EG), Wirksamkeitsmessungen (EF), Vitalparameter (VS)

Neben diesen drei generellen Domain-Kategorien gibt es zusätzliche Domains, u.a. für Demografie (DM) und Kommentare (CO).

3.2 Variablen

Jede Variable, die in einer Domain enthalten ist, wird wie folgt klassifiziert:

- Ursprung (aus welchem Datensatz kommt die Variable, ist sie berechnet oder zugewiesen)
- Funktion (dient die Variable zur Identifizierung oder Selektion, beschreibt sie Beginn/Ende/Dauer des Ereignisses, beschreibt sie das Untersuchungsergebnis)
- Eigenschaften (Label, Typ numerisch/character, enthält die Variable einen Code)

3 CDISC - SDTM

Eine der CDISC-Arbeitsgruppen hat das SDTM (Study Data Tabulation Model)-Modell entwickelt. Es basiert auf SDS und beschreibt Inhalt und Struktur klinischer Studiendaten, die einem Medizinischen Reviewer oder Auditor Einblick in die Patientendaten geben sollen. SDTM gibt genau vor, welche Datensätze (Domains) zu erstellen sind und welche Variablen in jeder Domain enthalten sein müssen. Es legt fest, wie die Variablen zu benennen und zu definieren sind.

Um SDTM-Datensätze zu erstellen, gibt es folgende Möglichkeiten:

- SDTM wird schon im klinischen Datenbanksystem (nicht auf SASTM-Basis) definiert
- SDTM wird erst nach dem Einlesen der Daten aus der klinischen Datenbank in die SASTM-Umgebung definiert
- eine Kombination aus den beiden Varianten: die sogenannte „Hybrid“-Lösung

3.1 SDTM innerhalb des klinischen Datenbanksystems

Die SDTM-Struktur wird schon innerhalb der klinischen Datenbank erstellt.

Die Vorteile dieses Lösungsansatzes sind:

- Rohdaten in der klinischen Datenbank entsprechen den SDTM-Daten
- einfache Kommunikation zwischen Datenmanager und Reviewer
- sobald die klinische Datenbank definiert ist, sind auch die SDTM-Datenstrukturen verfügbar

Nachteile dieses Lösungsansatzes sind:

- mehr Zeitaufwand, da alle Rohdaten in die entsprechende SDTM-Struktur überführt werden müssen
- abhängig vom klinischen Datenbanksystem kann diese Strategie umständlich sein und auch zu Qualitätsverlusten führen

3.2 SDTM innerhalb der SASTM-Umgebung

Die SDTM-Struktur wird erst nach dem Transfer der Daten aus der klinischen Datenbank in die SASTM-Umgebung definiert.

Die Vorteile dieses Lösungsansatzes sind:

- Änderungen an SDTM-Daten möglich, ohne den Datenmanagementprozess zu stören
- das klinische Datenbanksystem wird nur zu dem benutzt, was es am besten kann (Dateneingabe, Datenmanagement)
- große Flexibilität von SASTM erlaubt jede Transformation in eine SDTM-Struktur

Nachteile dieses Lösungsansatzes sind:

- Verzögerung der Prozesse: erst nach Aufsetzen der klinischen Datenbank können die SDTM-Datensätze erstellt werden
- Kommunikation zwischen Datenmanager und Reviewer durch unterschiedliche Datenstrukturen erschwert

3.3 „hybrid“-Lösungsansatz

Die SDTM-Struktur wird soweit wie möglich bereits in der klinischen Datenbank definiert („Pre-Processing“). Nach dem Transfer der Daten aus der klinischen Datenbank in die SASTM-Umgebung werden alle fehlenden SDTM-Strukturen mittels SASTM umgesetzt („Post-Processing“).

Die Vorteile dieses Lösungsansatzes sind:

- einfache Änderungen im klinischen Datenbanksystem sind leicht zu implementieren (Benennung von Datensätzen/Variablen, einfache Berechnungen)
- komplexere Anpassungen/Umrechnungen von SDTM-Variablen können innerhalb von SASTM automatisch durchgeführt werden

Nachteile dieses Lösungsansatzes sind:

- zusätzliche Validierung der in der SASTM - Umgebung durchgeführten Transformationen ist notwendig
- einige Zeitverzögerung der Prozesse

4 Implementierung von CDISC – SDTM bei Accovion

Bei Accovion haben wir uns für die „hybrid“-Lösung entschieden. Beim Aufsetzen der klinischen Datenbank (wahlweise CLINTRIALTM oder OracleClinicalTM) werden die Anforderungen aus SDTM soweit wie möglich für Dateien und Variablen bereits umgesetzt. Der Datenmanager ist dadurch mit den SDTM-Strukturen vertraut. Es werden zusätzlich „Komfortvariablen“ definiert, die den SAS-Programmierer bei der späteren Auswertung unterstützen (z.B. numerische Codes für Textfelder, numerische Datumsformate). Während des „Post-Processing“ in der SASTM-Umgebung werden zusätzliche Variablen z.B. aus der Kodierungstabelle MedDRA (Medical Dictionary for Regulatory Activities), die für Nebenwirkungen verwendet wird, zu den SDTM-Daten hinzugefügt. Zusätzlich werden alle Variablen, deren Transformation komplizierter ist, hier berechnet.

Zusätzlich zu den SDTM-Datensätze werden bei Accovion die Analyse-Datensätze gemäß der vorgegebenen CDISC ADaM Standards (Analysis Data Model) definiert, die auch für den Transfer der klinischen Daten an die amerikanische Zulassungsbehörde empfohlen sind. Hierin werden Auswertungskollektive und weitere Analysevariablen, die beim Aufsetzen der klinischen Datenbank noch nicht feststehen (wie z.B. Analyse-Zeitfenster, Baseline- und Endpoint-Definition), aufgenommen. Um finale SDTM-

Dateien zu erhalten, werden zuletzt die „Komfortvariablen“ wieder herausgelöscht, bevor sie an die Behörden transferiert werden.

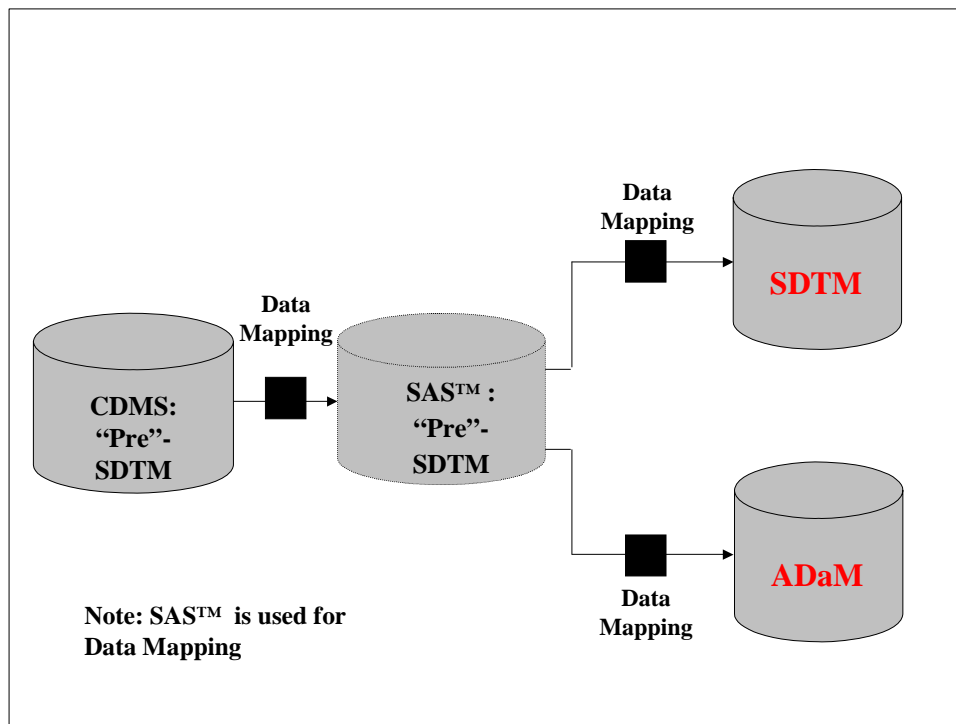


Abbildung 1: Implementierung SDTM

Wir haben uns für die „hybrid“-Lösung entschieden, weil sie die Nachteile der beiden anderen Lösungen ausschließt bzw. mindert. Die „hybrid“-Lösung erlaubt uns, die von der amerikanischen Zulassungsbehörde geforderte Standardisierung möglichst zeitnah umzusetzen, indem wir einen Teil der Standardisierung bereits im klinischen Datenbanksystem (ca. 70% der Variablen), den anderen Teil in der SAS™-Umgebung implementieren.

5 ADM-Macro

Das ADM (Analysis Dataset Model)-Macro wandelt die in der SAS™-Umgebung vorliegenden „Pre-SDTM“-Datensätze zu endgültigen SDTM-Dateien um. Das ADM generiert lauffähige SAS™-Programme. Werden diese vom ADM-Macro erzeugten SAS™-Programme gestartet, erstellen sie den gewünschten SDTM- (oder ADaM-) Datensatz.

Voraussetzungen für diesen Prozess sind:

- ein Excel™-sheet pro klinischer Studie („table of domains“), in dem alle zu erstellenden SDTM- (oder ADaM-) Datensätze abgelegt sind
- ein Excel™-sheet für jeden dieser Datensätze („table of mapping rules“) mit allen zu erstellenden Variablen

- eine Sammlung von SAS™ - Utility-Macros, z.B. zur Berechnung von Standardvariablen, die in mehreren Datensätzen benötigt werden

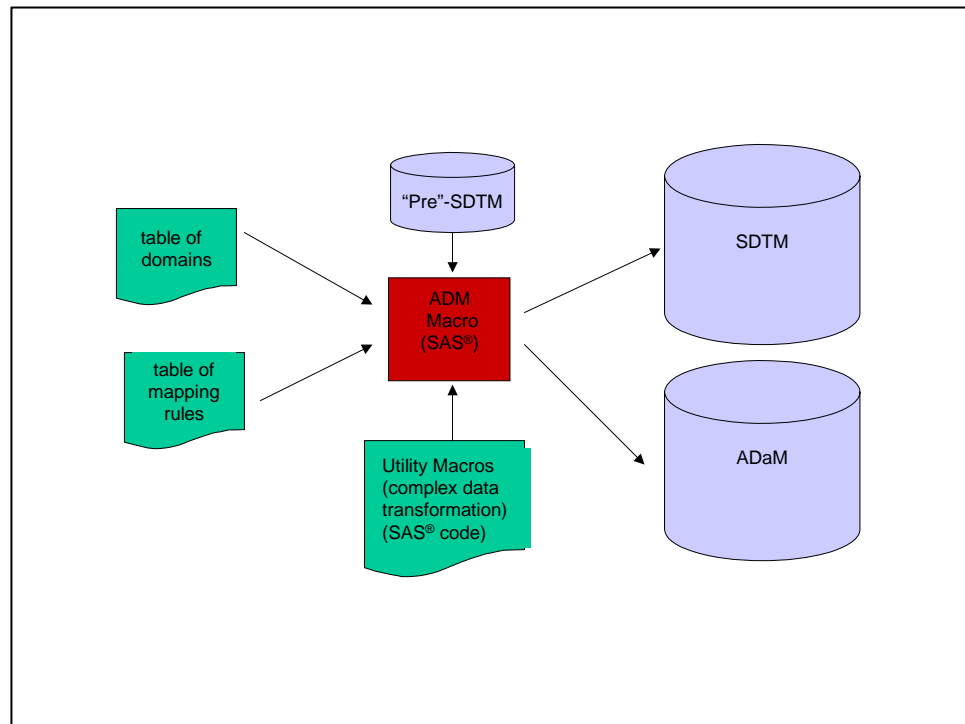


Abbildung 2: Implementierung ADM-Macro

5.1 „table of domains“

Das Excel™-sheet „table of domains“ wird einmal pro klinischer Studie erstellt. Es beinhaltet folgende Variablen: Studien-ID, Domainname (z.B. „AE“), Domain-Beschreibung (z.B. „Adverse events“), Domain-Kategorie (z.B. „E“ für „Event“) und die Sortierreihenfolge (z.B. „Subject-ID Startdatum Startzeit“).

Dieses Excel™-sheet

- legt die Reihenfolge der Abarbeitung für die SDTM-Datensätze fest
- legt die Sortierreihenfolge fest, nach der die Records im SDTM-Datensatz später sortiert sein sollen
- dient als Spezifikationsdokument, in dem der Kunde definiert, welche Datensätze mit welchen Variablen erstellt werden sollen

5.2 „table of mapping rules“

Das Excel™-sheet „table of mapping rules“ wird innerhalb einer klinischer Studie für jeden zu erstellenden Datensatz erstellt.

Es beinhaltet folgende Informationen:

- Studien-ID
- Domainname (z.B. „DM“)
- Variablenname (z.B. „AGEU“)
- Variablenbeschreibung (z.B. „Age unit“)
- Typ, Länge, Format der Variablen

- Reihenfolge der Erstellung
- Quelle der Variablen (Verzeichnis, Datensatz, ursprüngliche Variable)
- ob die Variable berechnet, zugewiesen oder übernommen wird
- SASTM-Code bzw. Macro-Aufruf.

Dieses ExcelTM-sheet

- definiert jede zu erstellende Variable genau
- legt die Reihenfolge der Erstellung für alle Variablen fest
- ruft die SASTM-Utility-Macros auf
- definiert den SASTM-Code, der bei der Berechnung der Variablen benutzt werden soll
- dient als Spezifikationsdokument, in dem der Kunde definieren kann, wie die Variablen eines Datensatzes erstellt werden sollen

6 Resümee

Der oben beschriebene Prozess, in dem aus ExcelTM-sheets SDTM-Dateien erstellt werden, zeichnet sich durch hohe Transparenz aus. Die ExcelTM-sheets dokumentieren genau, welche Datensätze mit welchen Variablen zu erstellen sind. Jederzeit ist ein Update der Datensätze durch Änderungen an den ExcelTM-sheets (und Neustart des ADM-Macros) möglich.

Die Erstellung der SDTM- (bzw. ADaM-)Datensätze

- wird durch die hohe Standardisierung beschleunigt
- ist durch die Benutzung des ADM-Macros und der SASTM - Utility-Macros weniger fehleranfällig
- ist mit geringerem Arbeits- und Trainingsaufwand zu erreichen
- ist mit weniger Validierungsaufwand zu erreichen, da standardisierte und validierte SASTM - Macros benutzt werden

Durch diesen Prozess wird außerdem die Integration von klinischen Studiendaten in eine bereits bestehende Datenbank bzw. die Generierung einer studienübergreifenden Datenbank erleichtert.

Referenzen

- [1] Mitteilung der FDA:” FDA Announces Standard Format That Drug Sponsors Can Use to Submit Human Drug Clinical Trial Data”
www.fda.gov/bbs/topics/news/2004/NEW01095.html

- [2] Submission Data Standards, Version 3.1
www.cdisc.org/models/sds/v3.1

- [3] Study Data Tabulation Model, Version 1.1
www.cdisc.org/models/sdtm/v1.1/index.html

- [4] Vortrag von Edelbert Arnold und Ulrike Plank, Accovion GmbH, auf der PHUSE- (Pharmaceutical Users Software Exchange) Tagung zum Thema „Customer oriented CDISC Implementation“ , Oktober 2005
www.lexjansen.com/phuse/2005/cd/cd10.pdf