

米国他における CDISC に関連する現状と将来展望

• 要約

ここ数年間、「CDISC standard」という名称は、誤って用いられてきた。CDISC (Clinical Data Interchange Standards Consortium) は 1990 年代後半の設立以来、数多くの標準(standard)を策定しているため、より正確には「CDISC standards」であろう。これらの標準の一部は、FDA が現在用いている臨床試験の集計データ提出用「規格 (specification)」として採用されている。このようなデータは FDA に市販申請を行う際に添付が求められる。FDA が行った規格公示の直接的結果として、CDISC 標準の一部はバイオ医薬品業界 (製薬会社、提携企業、業務受託機関等) に採用されている。他に、FDA への提出とは直接関係のないデータ演算 (the operational use of data) に用いられている CDISC 標準もある。規制当局での需要に加え、業界もデータの内容・フォーマットを標準化することの利点に注目し始めている。本文書では、業界と CDISC 標準との現在の関わり及び今後の見通しについて述べる。

• 緒言

「CDISC 標準 (CDISC standard)」という名称は、一般に CDISC (Clinical Data Interchange Standards Consortium) の作業チームのメンバーらが過去数年間に策定した個々の CDISC 標準の総称として、あるいは最も目に触れる機会の多い SDTM (Study Data Tabulation Model) を指して、多くの人々に用いられてきた。上述の作業チームを構成するメンバーは、米国 FDA 職員、製薬会社及びバイオテクノロジー企業社員、テクノロジー企業社員、CRO (Clinical Research Organization) 職員、他に業界のコンサルタント及び専門家らである。策定された標準は、臨床試験のみならず非臨床 (動物) 試験において収集、共有、保存、分析、報告されたデータの多様な使用法を示すものである。CDISC は、目標の一つとして、個々の標準を統一し、臨床試験のあらゆる段階のデータ管理、並びにこれらのデータの事後的な解析及び報告における使用を促進しなければならない。

前述の標準は公示され、業界において採用されたが、その歩調は多くの要因のため一定ではなかった。その要因として、企業が社内で何年もかけて開発した内部システム及び社内標準に満足していた事が挙げられる。既存のシステムや標準を変更する事に対する抵抗に加え、CDISC 標準が進化し続け、新しい版を重ねる度に増補された事も要因の一つである。

初期に CDISC 標準を採用した企業においては、新版の「旧版との互換性」が懸念された。本文書で探索する上記及びその他の要因が原因となって、CDISC 標準の採用が一定の歩調で進まなかったのである。

CDISC 標準採用の外的必要性というきっかけがなければ、CDISC 標準採用は未だ初期の段階であったであろう。これには、FDA 内部のイニシアチブの影響に加え、FDA が積極的に関わり、業界に対して CDISC 標準を採用するように動機付けを行う事が必要であった。本文書では、このような外部要因の幾つか及び FDA のイニシアチブに注目し、これらの要因がいかにか CDISC 標準に影響したか、また、CDISC 標準を採用する企業にいかにか影響しているのかという点に関して述べる。

さらに、本文書全体を通して CDISC 標準採用の様々な方法について述べ、各方法の利点及び欠点を考察する。

長年に渡り、バイオ医薬品業界の個々の企業において、他社とのデータ標準化に対する動機付けとなるものはほとんどなかった。このような企業の多くの意思決定者らは、データの収集、処理、及び FDA への報告について、自社の方法が他社の方法と同等か、あるいはそれ以上であると考えていた（FDA の審査担当の一部はそうは思わなかったかもしれないが）。

1990 年代を通して、FDA は各企業独自のコンピュータ・ハードウェア及びソフトウェア・システムを介したデータを受理した。これらのデータは、市販申請の正式書類、特に NDA（New Drug Application）に添付されて FDA の審査担当へ送付された。このような企業独自のシステムは、FDA に対して 2 つの重要な問題を引き起こした。

第一に、これらのシステムは多くの場合、同一の製薬会社によるものであっても異なっていた事である。このため、審査担当者の研修及びシステムのサポートは、製薬会社の手に委ねられていた。

第二に、これらのシステム及びシステム内のデータが、集計データの正式なペーパーコピーの代替として使用された場合、NDA の審査結果が、ペーパーコピーが使用された場合と異なる可能性があった事である。これらのシステム内のデータが紙媒体によって提出された内容と異なっていたという証拠はないが、2 種類のデータが異なるというリスクは依然として残されていた。多くの場合、試験の集計データの正式なペーパーコピーは、データの意義に関して結論を導く際に用いられるデータでなかったため、FDA はこのリスクについて非常に懸念していた。

1997年、FDAは「*Electronic Records; Electronic Signatures*」⁽¹⁾と題する新規則を公示した。これによりFDAは電子フォーマットでの文書及びデータを受理できるようになり、添付のペーパーコピーを「正式なコピー」とする必要がなくなったため、この規則は非常に重要なものとなった。

上述規則の公示から1年以内に、FDAのCDER（Center for Drug Evaluation and Research）及びCBER（Center for Biologics Evaluation and Review）はそれぞれ、電子文書及びデータがいかんして受理及び保管されるかという点に関するFDAの考えを業界と共有するためのガイダンス文書の草案を公示した。これらの文書の草案では、電子フォーマットのみによるNDA又はBLA（Biologics Licensing Application）のファイル及びフォルダのフォーマット及び構造について述べられていた。また、これらの文書では、Adobe PDFファイル及びSAS XPT（「移送ファイル」）が集計データの提出及びこれらのデータセットに対応する文書作成用の推奨ファイル形式として示唆されていた。これらのファイル形式が推奨されたのは、各企業独自のものではなく、「オープン」な標準であったためである。この標準は、その所有者のみに財政的な利益をもたらす企業独自の標準ではなく、この種の標準を支持するFDAの指令を満たすものであった。

1999年、前述のFDAの2センターは「*Providing Regulatory Submissions in Electronic Format — General Considerations*」⁽²⁾と題する共通のガイダンス文書を公示した。これには、PDF及びXPTのファイル形式についての記述があり、ファイル編成及び命名に関する詳細については各センターが作成する他のガイダンスにて述べるというものであった。

CDERは同時に、電子NDAにおけるファイル構造について述べた「*Providing Regulatory Submissions in Electronic Format — NDAs*」⁽³⁾と題する1999年の極めて重要なガイダンスを公示した。これには、CRT（Case Report Tabulation）データセクションについての詳細な記述が含まれていた。

CRTセクションの詳細な記述には、3つのファイル形式が含まれていた。1つ目は、BLANKCRF.PDFと命名されたCRF（Case Report Form）ファイルで、これには生データの収集場所を示す注記が付けられるとのことであった。2つ目は、DEFINE.PDFと称されるPDFファイルで、これには提出されるデータセットの詳細を示す一覧及び各データセットの表が含まれていた。詳細な内容は、変数名、記述、データの種別、使用されたコード及びデコード、データに対する他の重要情報に対する「コメント」等であった。最後にCDERは、データセットについて述べ、企業がデータセット作成時に使用する際の重要な考慮すべき事項を幾つか示した。

CDER ガイダンスの付録には、安全性に関連する 12 の「領域 (domain)」, すなわち, 「人口統計学」領域, 「有害事象」領域等, 類似した種類のデータを含むデータセットが例示されていた。

CBER は 1999 年末までに, 電子 BLA 提出に関する「*Providing Regulatory Submissions to the Center for Biologics Evaluation and Research (CBER) in Electronic Format — Biologics Marketing Applications*」⁽⁴⁾と題するガイダンスを公示した。これは CDER ガイダンスに極めて類似したものであった。

これらのガイダンス文書を公示した後, FDA は業界に対し, 以後は NDA や BLA への企業独自システムによるデータ添付を認めないこと, 並びに業界に同ガイダンスに従った電子フォーマットによる提出を義務付けることを伝えた。

- 「標準」に関する初のガイダンスに対する業界の反応

業界は企業独自の電子システムを放棄することを余儀なくされ, FDA が不明確に定義した「領域」において, PDF 文書及び XPT データセットという新たな現実へと歩を進めた。

FDA のガイダンス文書作成と同時期に, 業界内でも標準に関する検討が行われていた。

1997 年, 業界内のメンバーから成るグループが会議を開き, 業界内でのデータ共有のための標準について議論を行った。これらの議論は実効的な標準に関するものであった。OMG (Operational Management Group) 及び業界標準用語集 (Industry-standards Glossaries) が注目されるようになり, 1998 年までに DIA (Drug Information Association) に SIAC (Special Interest Area Committees) が組織された。これが現 CDISC の先駆けであった。

SIAC は当初, モデリング及び学術用語 (Nomenclature) のための 2 つの作業チームで構成された。DIA から分離した後, 上記 2 チームは直ぐに 5 チームとなった。学術用語チームは後に用語集 (Glossary) (又は専門用語 (Terminology)) チームとなり, モデリングチームは SDS (Submission Data Standards) チーム, ADaM (Analysis Data Model) チーム, ODM (Operational Data Management) チーム, LAB (Laboratory Data) チームの 4 チームに分割された。

CDISC の初期の会議では、標準を定義する代替的な方法について議論され、臨床研究に関する重要専門用語を定義するために用語集チームが設置されたが、実際のデータ標準の定義に向けた明確な進展はほとんどみられなかった。

CDISC Submission Metadata Model の初版⁽⁵⁾は、DIA の 1999 年の年報にてプレゼンテーションされた。この中では、FDA が 1999 年に公示したガイダンスは申請時に必要な文書についてはかなり詳細に示していたが、データ構成要素を体系化する方法に関しては説明が不十分であることが示唆された。同ガイドラインは、データを標準的なテクニカル・フォーマット（オープンな SAS V5 移送ファイル・フォーマット）で提出する事を義務付けるのみであったが、臨床データの内容及び構造、すなわちメタデータを記載する PDF ファイル（DEFINE.PDF）による提出の先例を示すものであった。プレゼンテーションでは、メタデータを提示する方法についてさらに詳細に示すことが提案された。

2000 年 4 月までに、David Christiansen と Wayne Kubick は CDISC Submission Metadata Model 第 1.1 版を発表した。本モデル及び第 2 版は、「FDA 審査用のデータ提出に適用可能な意義のある標準の確立」への基盤であり根幹的なアプローチとなった。XPT 領域ファイルのデータを示し、各変数を全体的なコンテキスト内に位置付けるために、メタデータ文書にはメタデータ（「データに関するデータ」）を記述する。

メタデータモデルの作成者らは直ぐに、メタデータモデルの概念を具体的なデータセットに適用する方法を示す領域モデルを開発するため、製薬会社数社からの志願者で構成されるチームを結成した。その後間もなく、FDA の担当者が複数加わった。これが、CDISC のデータモデリングチームである SDS (Submission Data Standards) チームの始まりであった。

このメタデータモデルは後に増補され、各 CRT 安全性データ領域を示すメタデータのスプレッドシートを収めた PDF ファイル集が加えられた。また、メタデータモデルにはメタデータ提示用の記述的スプレッドシートをフォーマットするための PDF ガイドが添付されていた。さらに、これらのスプレッドシートには、「最良実施例」を示す注記、すなわち各変数がいかに使用されるべきであるかという内容のコメントが添えられていた。また、各変数を常に報告すべきか否かについてそれぞれの変数が定義するように割り当てる「CDISC Core Variable」指定も含まれていた。

上述の版の価値は、「標準的な」領域と共にメタデータの例を示した点にあった。また、これによりどの変数が具体的にどの領域に属するのか、そして、どの変数があらゆる文書の提出において不要であるのかが明確に定義されていた。

同版は提出用データの体系化の例を明確に示したため、一部の企業はこれを NDA 用データ提出の基盤として用いた。残念ながら、当時は CDISC の組織は非常に小さく、この標準化への率先的な取り組みについて知っていた企業は極わずかであった。

• その他の CDISC チーム：ODM，ADaM，LABS

メタデータモデル及び SDS チームの最初の成功は、データ収集プロセス及び提出を支援する標準により関心を持っていた他チームの注目を集めた。これがきっかけとなり、直ちに ODM (Operational Data Modeling) チームが結成された。

ODM チームが結成されたのは 1999 年であり、同年に CDISC は CDM (Clinical Data Management) システムのベンダー団体を会議に招き、試験で得られた臨床データに互換性を持たせる新標準の可能性を討議した。同年の夏、2 社の CDM ベンダーが CDISC にそれぞれが開発したデータモデルを業界標準として支持することを検討させようと CDISC と交渉を行った。

これら 2 社 (Phase Forward 社、PHT 社) は、あらゆるデータ収集システムからデータを治験依頼者の中央データベースに移動させられる理想的なモデルをすでに開発しており、それぞれのモデルが業界標準として採用されることを何よりも望み、CDISC に期待を寄せた。両モデルとも XML (拡張マークアップ言語、Extensible Markup Language) 使用に基づいていた。XML はデータ及び文書を体系的に表す新しい技術標準で、技術ベンダー間、特に電子商取引においてサポートが拡充していた。

CDISC の主な運営原理の一つはベンダーに対して中立的であることだったため、CDISC の運営委員会のメンバーらは両社を他のベンダー、CRO、2 大 CDM ソフトウェアパッケージである Clintrial 及び Oracle Clinical の各開発企業が参加する新しいチームに招き入れた。

当局における 2 つの異なる必要性 (審査及び保管) に対処するために同時進行していた 2 モデルに取り組む中、CDISC 参加者らは臨床研究界が以下の 3 つの異なるタイプのデータで構成されている事に注目し始めた。

Data Sources : データが発生する。

Operational Data : データが収集され、整合性が再検討され、品質基準を満たすよう管理される。

Submission Data : 通常 Operational Data より抽出され、規制当局へ送付される。

データ提出の規則は FDA が定めるが、このような規則がデータ収集の必要性を満たすに十分ではないと思われたため、CDISC は以下の 2 分野において標準又はモデルを開発する必要があると考えた：データを発生場所から治験依頼者の内部データベースに移送するオペレーショナル・モデル、並びにデータを依頼者から FDA へ移送する提出モデル。

CDISC は 2000 年に非営利団体として正式に認められ、業界全体におけるデータ標準を定義する活動を推進するため、本格的にさらなる活動を展開し始めた。

直ちに新しい 2 つのモデリングチームによる活動が開始された。LAB チームは、臨床検査値のオペレーショナルデータ交換標準の定義に、ADaM (Analysis Data Modeling) チームは臨床データ解析用の標準提出モデルの定義にそれぞれ取り組んだ。

• FDA の前進：Patient Profile Viewer の「CRADA」

2001 年初頭、FDA は公開会議を開き、FDA 内で開発した CTOC (Cumulative Table of Contents) ビューワと称する電子 IND (Investigational New Drug) 申請書閲覧ソフトを発表した。この閲覧ソフトは潜在的資質を示したが、FDA はソフトウェア開発の中止を決定した。

この頃、FDA は「CRADA」(Cooperative Research and Development Agreement) と称するプロセスを正式に使用し始めた。CRADA はソフトウェア開発業者及び技術供給業者と共同で FDA が使用する予定の有用なソフトウェアを開発するという協定である。ソフトウェア開発業者にとっての利点は、業者が当該ソフトのライセンス、著作権、及び知的所有権を保持することが可能であった事である。このため開発業者は、ソフトをバイオ医薬品企業又は他の提携企業等の関係企業に販売することも可能である。

CRADA により開発されたソフトの一つが PPV (Patient Profile Viewer) である。2001 年 12 月, FDA は PPV 開発のために CRADA を結ぶ提携先を探している旨の通知を公示した。このソフトを開発する目的は, CRT データベースから直接患者プロファイルを作成し閲覧できるようにすることであった。FDA は PPD Informatics 社を選択し, 同社の市販ソフト「CrossGraphs」のモジュールを開発した。

このモジュールは, 領域で体系化されたデータセットの集合を開くようにデザインされており, (表のフォーマットで体系化された) データを「患者プロファイル」表示に変換するようデザインされていた。患者プロファイル表示は, FDA により「個々の被験者のために収集され, 時間で体系化された様々な種類の試験データ (例 複数領域のデータ) の表示」と定義されている。この体系化により, 異なる領域で同時に又は連続的に発現する様々な事象間の関連性が明確に示される。例えば, 患者に被験薬を投与し (「EXPOSURE」領域に記載) , その後直ぐに有害事象が発現した (「AE」領域に記載) 場合等である。

FDA によると, このソフトを目的通りに動作させるには, 標準化されたデータセット及びメタデータを入力する必要があるとのことであった。標準化されたデータセット及びメタデータを使用することで, 審査担当者が要する患者プロファイル作成のための準備量が減少し, 申請者が PDF で患者プロファイルを提出する必要性もなくなると考えられる。PDF 形式の患者プロファイルは, 常に必要とされるわけではないが, 審査部門によっては提出を求める場合がある。

• PPV に対する CDISC の反応 : SDTM 第 2.0 版

CDISC の SDS チームには FDA の代表も参加していたため, SDS チームは FDA が Patient Profile Viewer のために CRADA を締結した事がわかっていた。PPV を目的通りに動作させるには, データを一貫した方法で構成する必要があった。

2001 年 11 月までに, 旧版の機能を向上させた Submission Metadata Model 第 2.0 版が発表された。2001 年 12 月, これに伴う CDISC *Submission Data Domain Model v2.0* (SDDM) ⁽⁶⁾ が発表された。この文書は, 全ての領域スプレッドシート例を 1 つの PDF 文書にまとめたものである。さらに, 仮定の一覧が追加され, 他の明瞭化点及び強化点に加え, 予測されるデータに関連するさらなる情報も示された。

第 2.0 版における最も顕著な変更点は、治験依頼者に「ECG」及び「Vital Signs」領域の「垂直 (vertical)」データセット及び「より正規化された (more-normalized)」データセットのいずれを提出するかという選択枝が与えられた事であった。これらの領域の正規化によって、FDA は「新しいデータセット及びデータ閲覧技術を導く」能力が得られると考えられる。また、審査のためのデータ保存、検索、他のデータとの併合がより柔軟に行えるようになると思われる。

垂直モデルの開発は、ECG 領域及び Vital Signs 領域の水平又は正規化の程度が低いモデルの改善にもつながった。第 2.0 版では、臨床検査値、ECG 及びバイタルサインの測定値用の標準化された LOINC (Logical Observations, Identifiers, Names and Codes) コードも導入された。

1990 年～2000 年の期間後半における市場のピーク時であったこの時期に、数多くの大きな製薬会社が医薬品の開発及び上市を目的に、より小規模な新興企業、あるいはバイオテクノロジー企業と提携し始めた。もう一つの傾向として、より小さな企業が乗っ取りや合併のターゲット (あるいは仕掛け人) になった。このような業務提携や合併は、業界標準が有益である可能性を示していたが、少数の企業は未だ CDISC 標準を採用していなかった。

SDDM 第 2.0 版が発表されると、業界は SDDM を以前より高く評価するようになり、CDISC の治験依頼者の多くが、この標準を提出準備プロセスに適用し始めた。自社の Global Data Integration Database (多くの場合 SAS において) に SDDM を適用し始めた企業さえあった。このようなデータベースは、Submission DB, Analysis DB, Integration DB, Global Integration and Analysis Databases (GIADB), あるいは単に「データ・ウェアハウス」と称される。

2002 年、Patient Profile Viewer を開発するために CRADA を締結した後、FDA は治験依頼者に PPV に対応する標準化された構造のデータを提出させる必要がある事に気付いた。CRADA プロジェクトは、データ提出のための一連の PPP (Patient Profile Pilot) 規格を公示した。FDA は Patient Profile Viewer をテストするため、CDISC の SDS チームに PPP 規格に従ったデータの提出を依頼した。また、SDDM 第 2.0 版標準及び PPP 規格の互換性の検討も行われる予定であった。

2002 年の夏、CDISC の SDS チームは SDDM 第 2.1 版の発表に備え、第 2.0 版の発表で得られたコメントを見直し、新版で達成すべき事柄を検討し始めた。SDS チームは、新たな領域、コード使用の拡張、既存領域と先に発表されていた ODM メッセージ・フォーマット標準との互換性の向上に着目していた。また、PPP 規格との互換性を高めるために SDDM

標準を修正する可能性も考慮された。その結果、CDISC SDDM 標準の新版には Patient Profile Pilot のフィードバックが組み込まれることが決定した。

また、SDS チームは CDISC のイニシアチブを広げ、米国内の標準を定める組織である ANSI (American National Standards Institute) に加盟している HL7 (Health-Level 7) 組織を通して CDISC 標準を公示することを検討していた。HL7 は、ヘルスケア及び医療提供者への償還の領域におけるデータを標準化するために開発されたものである。

8 月に行われた SDS チームの対面会議で、チームに参加していた FDA の代表者らは、FDA が Patient Profile Viewer 規格をデータ標準として発表するという計画があることを伝えた。FDA は、米国内の標準を定める組織である ANSI (American National Standards Institute) に加盟している HL7 (Health-Level 7) 組織を通してこの規格を公示することを提案した。HL7 は、ヘルスケア及び医療提供者への償還の領域におけるデータを標準化するために開発されたものである。このため、同目的の標準が重複する、あるいは競合するのではないかという懸念が生じたのももっともであった。

この会議では、FDA の他のニーズについても SDS チームに伝えられた。FDA は、提出された標準化データの全てを保存するデータ・ウェアハウスについて計画中であった。このウェアハウスによって、FDA がデータ、特に安全性の懸念を示すデータを「掘り出す」ことが可能なソフトウェアを製作できるようになることが期待されていた。また、データ・ウェアハウスにより、FDA が類似した医薬品又は医薬品の分類についてのデータを集積し、この集積されたより大きなデータを解析することが可能になるという利点が予想されていた。

上述の会議では、最も差し迫った問題は、第 2.0 版の問題点を是正し、8 月末までに第 2.1 版を発表することであるということによって一致した。会議後の連絡で、FDA が、HL7 により発表され「規範的な」標準として作成された CRT 標準を有することを望んでいる事が明らかになった。これにより、FDA はこの「標準」を参照できるようになり、PPV 規格を標準として公示する必要がなくなると考えられた。

CDISC の SDS チームは、第 2.1 版の修正点を発表後（実現のためではなく素案として）、次の第 3.0 版に取り組めるように、FDA が HL7 標準を求めていることに気付いた。しかし、業界の多くの企業は第 2.0 版及び第 2.1 版を採用しており、業界からは全く異なったものになると思われる版を採用することに対する抵抗が予想された。

LAB チームはこの時点で、中央検査施設からどのデータを発信すべきかを記述する LAB 標準の第 1.0.0 版を発表した。また、同チームはこのような LAB メッセージを体系化する XML メッセージの標準も作成した。

FDA はまた、安全性解析用のソフトウェア、データ・ウェアハウス機構、及びデータのバリデーション・ツール開発のため、CRADA を結ぶ提携先を探していた。

2003 年 3 月までに、SDS チームは「*Version 3.0 Submission Data Standards - Review Version 1.0*」⁽⁷⁾と題する新文書を作成した。これには、緒言、General Study Data Information Model、及び CDISC Submission Domain Model が含まれていた。第 3 版には第 2 版からの変更点が多かったが、目標は業界内の SDDM 第 2.0 版を採用していた企業に互換性を提供することにあった。これは、投票により正式な規範的標準となると考えられる最終版に先立って、HL7 組織が読み、コメントを加えるためのものであった。

HL7 のメンバーらがコメントを出し、重要なコメントに対処するための変更が加えられた後に、SDS チームは投票で HL7 により規範的標準と認められた CDISC の *Submission Data Domain Models Version 3.0 - Final Version 1.2*⁽⁸⁾用の文言に到達した。

互換性における例外の一つは、SDDM 第 2 版における水平（又は非正規化）領域が第 3.0 版では作成できない事であった。これは、最も重要な目標の一つが、FDA が CRT（垂直又は正規化）表示を「一覧」又は水平表示に変換する標準ツールを使用可能にする標準を提供することであったためである。非正規化データを正規化データに変換するツールを作成することは可能であったが、この作業はより困難かつ複雑であると思われた。このため、正規化データ提出を標準とする方が良いと判断された。

正規化データのもう一つの利点は、標準遵守のバリデーション及びデータのウェアハウスへのインポートが容易である事であった。

第 3 版の他の特徴は、第 1 版、第 2 版で元々記述された安全性領域以外の領域におけるデータのガイダンスが提供される事である。第 3 版では、各領域はあらゆる種類のデータに対し開かれていた。体系化を行うために、各領域は「所見 (findings)」、「介入 (interventions)」、あるいは「特殊目的 (special purpose)」に分類された。「特殊目的」領域は、第 3 版内で明確に定義されていたため、新しい領域は他の 3 領域のいずれかに分類される必要があった。

これらの領域の明確な文書化のために、新標準の作成が決定された。2003年4月、「DEFINE.XML」規格を作成するために、ODM、ADaM、SDSの各チームのメンバーで構成されるフォーカス・グループが結成された。また、この規格の要件及び利点を記述した白書が書かれた。この規格は、ODM XML 構造及びフォーマットを用い、従来は DEFINE.PDF 文書ファイルで提出されていたデータセット、領域、及び変数の情報を適用するようにデザインされていた。新規格では標準化された構造及びフォーマットを有する機械読取可能なファイルとなるため、FDA がデータを FDA のウェアハウスにロードできるという利点があった。CRT-DDS (Case Report Tabulation Data Description Specification⁽¹²⁾) が公示され、HL7 に提出された。この標準は、規格の提出、HL7 メンバーのコメントの受領、改訂版の再提出という HL7 の投票サイクルを何度も通過した。(CRT-DDA 第 1.0.0 版は最終的に 2005 年 2 月に正式に認められた。)

第 3.0 版を検討する 2 回目のパイロット試験が計画された。このパイロット試験は 2003 年末に予定された。結果は、2003 年 10 月に開催される FDA の公開会議で提示されることになった。パイロット試験には 8 社が参加し、第 2 版フォーマットのデータを提供した。1 社は旧版のデータを第 3.0 版の提出用垂直領域へとマッピングした。第 3.0 版は、データをウェアハウスにインポートするよう機能可能であると結論付けられた。また、提出用領域の作成をより明確に行えるように第 3.0 版の機能を向上させる必要があるとの結論も導かれた。

また、業界に対して提出準備プロセスに適用する標準として、第 3.0 版を推奨するのは時期尚早であるとの見解の一致がみられた。これは勧告であったにもかかわらず、第 3.0 版の適用を試みた企業が数社あった。しかし、ほとんどの企業は第 2.0 版又は第 2.1 版に準拠した文書提出を継続した (第 2.1 版はドラフト版として発表されていただけであったが) 。

FDA のパイロット試験担当者、パイロット試験参加者、CDISC からのコメントのレビュー後、SDS チームは 2004 年 6 月までに 2 つの文書 (*CDISC Submission Data Tabulation Model version 1.0 (SDTM)*⁽⁹⁾ 及び *SDTM Implementation Guide V3.1 (I.G.)*⁽¹⁰⁾) で構成される新版を作成していた。

パイロット試験で得られた教訓は、SDTM IG 第 3.1 版の付録として公示された。以下に抜粋を示す。「本パイロット試験で得られた第一の教訓は、不整合を減らし、これらのモデルの理解を深めるためには、ガイダンス及び規格の追加が必要である。具体的には、規格、規則をより明確に伝え、例を用いたガイダンスを追加するために、詳細な実施ガイドが必要である。また、本チームはデータセットの垂直性により、最高レベルのメタデータ

(例 define.xml を介して)を提供するための管理された特有の専門用語の重要性及び必要性が強調されることがわかった」

明確さを増すため、SDTM IG 第 3.1 版にはより多くの例が含まれていた。また、この版では「試験デザイン (Trial Design)」用領域が導入される。このような領域は、試験で用いられる複数の群について (群の構成要素及び各群が全体とどのように関連するかを定義することによって) 記述する。また、被験者がどのように検討されるのかという点についても記述する (例えば、クロスオーバー試験においてどの群を追跡しているのか等)。被験者らが計画通り試験を受けられるようにするため、これらの領域及び他の被験者データ領域が比較されることも考えられる。

2004 年の 8 月から 12 月にかけて、ADaM チームは「*Statistical Analysis Dataset Model: General Considerations Version 1.0*」⁽¹¹⁾の 5 草案及び最終版を発表し、一般構造、メタデータ、並びに解析データセットにおいて通常みられる内容を示した。このガイダンスは、CDISC *Submission Metadata Model* に従い、また、機械読み取り可能フォーマットでの解析メタデータ提出用の機構として「Define.XML」(CRT-DDA 第 1.0 版)を参照して、SDTM 第 3.1 版の用語 (nomenclature) に基づいて作成された。

• eCTD 規格における SDTM 第 3.1 版に対する FDA の認識

eCTD 規格は 2003 年 11 月、ステップ 5 (実現) に移行した。FDA は 2004 年 3 月 14 日、ホームページ上に eCTD に対する FDA の見解を掲載した。

2004 年 7 月、FDA は eCTD の適用に必要な eCTD ガイダンスへの補足規格である *Study Data Specifications v1.0*⁽¹³⁾を公示した。Study Data Specifications のこの版では、FDA に eCTD 構造で Data Tabulation データセットを提出する際に従うべき標準として CDISC SDTM 第 3.1 版が参照された。

SDTM 第 3.1 版が参照されたことで、CDISC 組織及び SDTM 標準の認知度は顕著に上昇した。FDA は SDTM 第 3.1 版を参照することにより、eCTD フォーマットに含まれる予定の試験データ集計に携わる者全てに対し、FDA が CDISC の重要性を認識していること伝えたのである。また、これにより SDTM 第 3.1 版標準が将来の文書提出用に採用されることになることにより強く業界に印象付けられた。

eCTD データ規格が、FDA が CDISC SDTM 第 3.1 版を選択したことを示したにもかかわらず、2004 年の時点では、多くの企業ではまだ eCTD フォーマットを用いて文書提出を行う準備が整っていなかった。実際、FDA の報告によると、2004 年度に CDER 及び CBER に eCTD フォーマットで提出されたのは、市販申請（NDA 及び BLA）12 件、IND2 件、添付文書提出 100 件超であった。これは、2004 年に報告された計 137 件の原本による（original）市販申請及び 81 件の市販申請書再提出と大差のない数字である。企業は 1999 年の電子提出ガイダンスを用いた申請書の提出に慣れてしまっており、eCTD の採用に向けて動き出そうとする動機が欠けているようであった。FDA は、eCTD の使用を提出用の必須フォーマットとしてまだ義務付けていなかった。企業には電子提出を行う際に、1999 年ガイダンスのフォーマット及び eCTD フォーマットという 2 つの選択枝が与えられていたのである。

他に FDA において進められたのは、FDA に臨床データを提出する際は標準化された電子フォーマットを使用する事を義務付けようとするイニシアチブであった。2003 年 9 月、FDA は官報で NPRM（Notice of Proposed Rulemaking）を公示した。この通知は、標準化された電子フォーマットによるデータ提出を義務付けるための規則改正に向けた第一段階の一環であった。NPRM は 2005 年 6 月という実効日程の提案を伴い、2004 年 12 月に再公示された。2005 年 5 月に出された官報の議題一覧の中で FDA は、実効日程を 2005 年 10 月に延長した NPRM を再公示した。2005 年 10 月、保健社会福祉省（HHS：FDA が属する省）は官報の HHS 規制計画で 2006 年の同省の優先事項を公示し、上位 7 項目の一つとして標準化された臨床データの提出を挙げた。NPRM の細目では、規則改正実施の時間枠は 2 年とされていた。

NPRM ではさらに、データレビューをより効率的にし、紙媒体のデータを手作業で FDA の電子システムに入力する場合に起こり得る誤りを少なくするデータを義務付ける理由が言及されている。より効率的なデータ処理及びレビューの他に、標準化の利益としてデータをより効率的に保管可能になることが述べられている。

また、FDA は 2005 年 5 月、改訂 eCTD *Study Data Specifications* 第 1.1 版⁽¹⁴⁾を公示した。これらの改訂規格では、臨床試験データ標準のために CDISC SDTM 第 3.1 版が引き続き参照されていたが、他の CDISC 標準も加えられていた。

集計データについては、新 eCTD データ規格は、Clinical SDTM 第 3.1 版モデルに適合するよう作成されたが、動物の毒性データ規格に適用された CDISC の SEND（Standard for Exchange of Nonclinical Data）が参照された。この標準は、2002 年に結成され、SDTM と並行して本ガイドラインを作成した CDISC の SEND チームにより作成された。SEND チームは 2004 年 12 月に最新の第 1.7.5 版⁽¹⁴⁾を、2005 年 3 月に実施ガイドをそれぞれ公表した。

文書化（データ定義ファイル）については、新 eCTD データ規格では CDISC CRT-DDS（define.xml）が、eCTD 提出においてメタデータを提示する望ましい方法として参照されている。

解析データについては、唯一公表されているガイドラインが、包括的な規格文書として用いるには具体性が十分ではないとみなされていた一般的考察文書であったため、この eCTD データ規格ではまだ CDISC は参照されなかった。

● 業界における標準の実践状況

1999 年ガイダンスの提出用標準が eCTD 標準に切り替わる事が発表された事に加え、臨床試験データの電子媒体での提出を義務付ける前述の規則改正が提案された事によって刺激された業界は 2005 年、CDISC SDTM 標準を採用する必要性を真剣に捉え始めた。eCTDs による提出は増加し、2005 年 9 月時点では 2003 年～2005 年の提出件数は NDA で 46 件（提出総数は 588 件）、BLA で 11 件（提出総数は 233 件）、IND で 43 件（提出総数は 234 件）であった。（2005 年度の最終的な総計は 2006 年 6 月に報告予定）

eCTD 提出の増加は、CDISC 標準の実施促進活動の高まりに平行してみられる。2003 年から 2005 年初旬にかけて、既存の社内標準を用いていた多くの企業は、標準の転換方策を立案し、提出データ作成時に CDISC SDTM データ標準へのマッピングを行った。

企業は、この方策を実行するという事は試験のデザイン時、症例報告書のデザイン時、あるいはデータ収集時に CDISC 標準が要求又は期待するデータが考慮されていない可能性があるというリスクを冒す事になるとの感を持った。これが本当であれば、CDISC 標準が要求又は期待するデータが欠失し、試験に悪影響を及ぼす可能性があった。

多くの企業は、CRF データを収集及び検証する DMS（Data Management Systems）内で「CDISC 様の（CDISC-like）」又は「CDISC に適合する（CDISC-friendly）」変数の使用を検討し始めた。これらのデータ変数は、CDISC SDTM 第 3.1 版標準のサブセットとなり、DMS 内の SDTM 第 3.1 版変数名を用いることになっていた。CDISC に適合する変数の他の例は、収集されたデータを表す固有変数であったが、データを真の SDTM 変数にするには、データにある種の公式を適用する必要があった。これらの変数は、SDMT 第 3.1 版の変数

名と類似した名称を有する可能性があったが、真の SDTM 第 3.1 版変数との明確な相違点を有する場合もあった。

例えば、SDTM 変数である「年齢 (AGE)」は、症例報告書 (Case Report Form) に記入されることがほとんどない。「生年月日 (date of birth)」及び「無作為化実施日 (randomization date)」は記入されるため、「年齢」は通常、減法により求められる。多くの企業では、このような変数の算出の多くはデータ管理担当課ではなく、生物統計プログラミング担当課が行っている。このようなことは、確実に適切なアルゴリズムが一貫して用いられ、データが適正に解析に供せられるようにするために多くの企業で行われてきた。

このような責任の分割はさらに、提出用に作られた SDTM 第 3.1 版フォーマットでデータを作成する際に、また、解析用データセットを作成するのにどのデータを原データとして用いるかという問題に関わってくる。Susan J. Kenny 及び Jack Shostak が PharmaSUG (Pharmaceutical SAS User Group) の論文^{(15),(16)}の中で上述のプロセスへの対処法を幾つか提案している。

Shostak 氏が提案⁽¹⁵⁾した SDTM 変数を作成する際の 3 つの対処法を以下に示す。

- 1) SDTM の構築を完全に DMS 内で行う。(初期段階での作成)
- 2) SDTM の構築を完全に SAS 内で行う。(最終段階での作成)
- 3) 2 法を組み合わせた複合型の方法で SDTM の構築を行う。(初期段階+最終段階)

SDTM の実践に際し、初期の頃は多くの場合、柔軟性を欠く、あるいは企業独自の DMS 構造から SDTM 構造へのデータのマッピングが柔軟に行える最終段階での作成 (方法 2) が用いられた。また、これにより SDTM 変数は DMS 外に保存された。SDTM モデルの第 2 版から第 3.1 版への改訂が急速に行われていたため、この点が重要視されていた。しかし、現在では第 3.1 版が定着し広く採用されているため、それ程重要視されていない。

初期段階での作成 (方法 1) の最大の欠点は、DMS 内では作成されなかった追加的な変数を DMS 内で作成しなければならないという事である。このことは、これらのシステムに対する余分な労力と費用がかかることを意味している。さらに、従来は生物統計プログラミング担当課が作成していた上述の追加的な変数の結果を受け入れる事に対し、同課が難色を示す可能性がある。

複合型の対処法 (方法 3) は、最も一般的に受け入れられている方法になりつつある。この方法では、従来は DMS で収集、検証、保存が行われていた「生の」、つまり収集された

変数に対し、DMS で「CDISC に適合する」命名規則が用いられる。次に、データがエクスポートされ、通常は SAS を用いて残りの CDISC SDTM 変数及び解析データセットが生成される。

CDISC 変数の一部は、有効な計画を立てないと作成することができない。一例として、「USUBJID」の変数名を持つ「Unique Subject Identifier」用の SDTM 変数が挙げられる。この変数は、薬剤開発プログラムに参加する各個人に固有のものとなる。プログラムに参加する被験者らが 1 件の試験に参加後、後続の試験に進む場合、USUBJID を DMS で作成するために、USUBJID を一意のものにする方式は一連の試験開始前に決定する必要がある。

この変数の複合型対処法は、DMS で SUBJID 等の「CDISC に適合する」変数を使用することである。この変数を使用する事は、この被験者識別子が試験レベルでのみ一意であることを示すと思われる。この変数はエクスポートされ、SAS において SUBJID をプログラムレベルで USUBJID にマッピングするのに使用される。この USUBJID は、あらゆる解析又は提出準備の整った SDTM データセットのエクスポートに使用される。

上述の複合型対処法でもやはり、DMS の構築を行っているデータ管理担当者、あるいは SAS で作業を行う生物統計プログラマー同士の意思疎通及び合意を要する。誰が各変数の作成の責任者であるのか、CDISC に適合するどの変数が DMS から直接移されるのか、また、他の最終 SDTM 第 3.1 版の変数が作成するのにどれが用いられるのかについて事前に合意に達していなければならない。

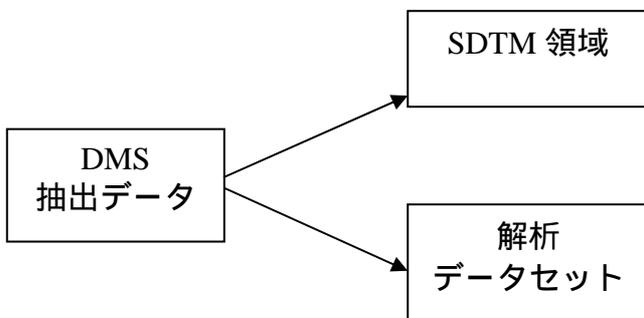
● 解析準備

2005 年 4 月、ADaM は Change from Baseline Analyses , Categorical Data Analyses , 及び Subject-Level Analyses 用の 3 つの規格草案を発表した。これらの解析（及びその他の解析）を準備するために、Susan Kenny は以下の 4 つの方法のいずれかを用いることを提案している⁽¹⁶⁾。

- 1) 平行法 (Parallel Method)
- 2) 回顧的方法 (Retrospective Method)
- 3) 線形法 (Linear Method)
- 4) 複合法 (Hybrid Method)

以上の方法は、DMS、解析データセット、及び SDTM データ領域間の関係を定義するようデザインされている。

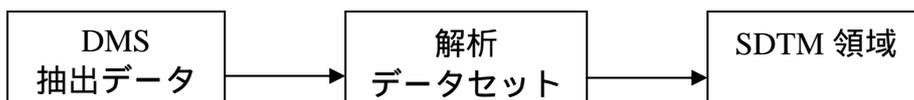
平行法（方法 1）について以下に図示する。



この方法では、DMS からエクスポートされたデータが原データとして使用され、SDTM 及び解析データが作成される。しかし、これら 2 種のデータは互いから別々に生成される。これにより、作業を 2 チームに分けて行うことが可能になる（SDTM 処理チーム及び解析処理チーム）。これら 2 チームは社内に設置しても良いし、作業を外注しても良い。

この方法の最大の欠点は、解析データでは SDTM 変数が原データとして使用されない事である。SDTM データ又はデータセットのみしか受理しない FDA の統計担当者らが解析を再現しようとした場合、困難が伴うと思われる。さらに平行法では、提出される 2 種類のデータ間の一貫性を保つために、2 チーム間の密接な意思疎通及び合意が必要となる。データの不整合があると FDA の審査担当者から重大な質疑を受けることになり、承認が遅れる可能性がある。

回顧的方法（方法 2）について以下に図示する。



この方法では、DMS からエクスポートされたデータが解析用の原データとして使用される。解析が完了した後、SDTM データが作成される。

最大の利点は、解析結果が試験の不成功を示し提出が行われない場合、SDTM データを生成する必要がない事である。

この方法には多くの欠点がある。平行法と同様に、FDA の統計担当者は解析用の原データが得られない。DMS が CDISC に適合していなければ、解析に使用し SDTM 領域に転送する変数は、解析用に SDTM 変数に変換する必要がある。SDTM が収集時の元データを表すように、補完された日付又は解析で行われたその他の種類のコーディングを取り消す必要がある。この方法は非常に非効率的であると思われる。

線形法（方法 3）について以下に図示する。



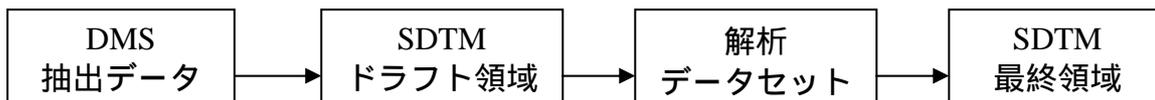
この方法では、SDTM の準備のために DMS からエクスポートされたデータを原データとして使用する。SDTM は解析用の原データとして使用される。

この方法は、最良の方法の一つであると思われる。DMS が CDISC に適合していない場合、DMS を SDTM 領域に変換するのに大きな労力を要する。SDTM データ領域が完成するまで解析が始められないので、この作業が全体的なプロセスを遅延させる可能性がある。さらに SDTM は、解析結果が肯定的な結果を示さない場合でも、あらゆる試験に適用される。

最大の利点は、FDA の審査担当者が解析用の原データを得ることができ、提出時に提供されたものと同じプログラム及びメタデータを用いて結果を再現できる事である。

この方法では、特にデータ管理又は解析の一部であっても外注される場合は、データ管理担当と生物統計担当との間に明確な意思疎通が必要である事は明白である。

SDTM 及び解析準備の複合法（方法 4）について以下に図示する。



この方法では、データは DMS からエクスポートされ、解析データセット用の原データとしてドラフト SDTM データが作成される。解析が完了した後、SDTM 提出領域が完成する。

この方法は、DMS が CDISC に適合していないことを前提としているようである。DMS から抽出されたデータは、必要に応じてのみ変換され、解析用原データとして適切な SDTM 領域データが作成される。解析結果により当該プログラムが提出されることが確認されると、最終 SDTM が作成される。

この方法の利点及び欠点は、線形法と非常に類似している。明らかな利点は、最終 SDTM 領域を全ての試験について作成する必要がなく、解析の結果、提出が適切であると判断されたプログラムについてのみ作成すればよい事である。

以上の 4 つの方法により、SDTM 及び解析データ準備のために業界が採用しているプロセスへの有益な洞察が得られる。さらに、SDTM 変数又は SDTM に適合した変数をデータ収集プロセスのできるだけ早い段階において適用する事の利点が示されている。

• 次の段階

CDISC は HL7 及び FDA 等のパートナーと共に、データ消費者がデータをどのように使用するのかを見出すための取り組みを行ってきた。本文書全体を通して述べたように、FDA は受理の効率化、審査の改善、データアーカイブ又はデータウェアハウス内でのデータ再利用のために、電子データの標準化に着目している。また、CDISC は他のデータ消費者にも注目しており、CDISC のニーズを満たす標準の定義に向けたイニシアチブが推進されている。

HL7 及び CDISC は、機械読み取り可能なプロトコルを定義するため、PR (Protocol Representation) プロジェクトに取り組んできた。SDTM の試験デザイン・コンポーネントは、PR グループのサブグループでもある。WHO、米国 NCI (National Cancer Institute)、及び他の試験データ登録機関等、一部のデータ消費者が特定されている。これらの機関は、被験者候補らに試験参加により得られる可能性のある利益について情報を与えられるよう

に、試験の詳細について把握することを望んでいる。試験の実施者らは、特にまれな疾患の治療法を研究する場合、このような機関を被験者に対する「宣伝」として利用することが可能である。

PR データの他の消費者は、試験の実施及び試験データの解析を行う者であろう。試験プロトコール中の文言が複数の意味に取れる場合が多い。このようなプロトコール中の曖昧な文章は、試験の実施時に問題となり、データ解析を特に困難なものにする可能性がある。明確で一義的な文言、あるいは機械読み取り可能なプロトコールは、このような問題を回避する上で有用であると考えられる。

CDISC 及び FDA が実施中のパイロット・プロジェクトでは、実際の試験で得られたデータが用いられている。このプロジェクトの目的は、FDA の審査担当者への SDTM 第 3.1 版データセット、ADaM 解析データセット及びメタデータ(注釈付き CRF、並びに DEFINE.PDF と DEFINE.XML の双方あるいはいずれかを伴う)の提出を確立することである。これによる審査からは業界に対するフィードバックが生まれる。パイロット・チームの目的は、このようなフィードバックが将来における提出を改善する一助となるよう、白書を著して得られた所見を提示することである。

もう一つのプロジェクトは、複数の CDISC 標準の調和を図り、単一の標準に統合することである。これは複数年を要するプロジェクトで、BRIDG プロジェクトと称される HL7 及び CDISC の UML (Unified Modeling Language) モデリングにおける試みに基づいている。BRIDG プロジェクトは、臨床試験のプロトコール・デザインから試験の実施及び審査に至るまでのあらゆる業務プロセスをモデリングするものである。このようなモデルは、複数の CDISC モデルを調和し、統合するための基盤となると考えられる。

LAB 及び ODM 等、他の標準は時折アップデートされている。LAB モデルは、治験依頼者及び中央検査施設間でデータを共有する独自のデザインを有するが、ODM モデルにはさらなる可能性がある。ODM XML モデルは CRT-DDS における DEFINE.XML の基礎であり、最終的には FDA への提出におけるデータセットのフォーマットになると思われる。ソフトウェア開発の出資者らが、データ保管又はデータ移送用に ODM を使用する可能性がある。SAS、Phase Forward (ClinTrial)、Oracle Clinical 等、より多くのソフトウェア・ベンダーが「近い将来」ODM をインポート及びエクスポートすることに関して議論を行っている。

• 結論

CDISC 標準は長年に渡って築かれてきた。しかし、業界は米国の規制当局がこのフォーマットによるデータを望むとの見解を公にするまで、CDISC 標準を受け入れず、採用しなかった。FDA がこのフォーマットによるデータを望んでいるという事に加え、間もなく他のいかなるフォーマットによるデータも受理しなくなると発表したため、業界は現在 CDISC 標準の採用に向けて急速に動いている。

業界は CDISC 標準の利点を理解しつつあり、データに関して議論しデータを共有するための共通構造について学んでいる。このような意思疎通は製薬会社及び FDA 間で行われる場合もあれば、提携企業、研究者、及び業務受託機関間で行われる場合もある。意思疎通を行うのが何者であれ、CDISC 標準は言語となり、データの収集、共有、保存、解析、報告、再利用を改善するはずである。

•参考文献

- (1) FDA - Electronic Records; Electronic Signatures (ERES rule)
- (2) FDA - CDER/CBER - Providing Regulatory Submissions in Electronic Format — General Considerations
- (3) FDA - CDER - Providing Regulatory Submissions in Electronic Format — NDAs
- (4) FDA - CBER - Providing Regulatory Submissions to the Center for Biologics Evaluation and Research (CBER) in Electronic Format — Biologics Marketing Applications (BLAs)
- (5) CDISC Submission Metadata Model
- (6) CDISC Submission Data Domain Model v2.0 (SDDM)
- (7) CDISC Version 3.0 Submission Data Standards - Review Version 1.0
- (8) CDISC Submission Data Domain Models Version 3.0 - Final Version 1.2
- (9) CDISC Submission Data Tabulation Model version 1.0 (SDTM)
- (10) CDISC SDTM Implementation Guide V3.1 (I.G.)
- (11) CDISC Statistical Analysis Dataset Model: General Considerations Version 1.0
- (12) CDISC Case Report Tabulation Data Description Specification (CRT-DDS) v1.0.0
- (13) FDA - eCTD - Study Data Specifications v1.0
- (14) FDA - eCTD Study Data Specifications v1.1
- (15) CDISC Standard for Exchange of Nonclinical Data (SEND) v1.7.5
- (16) PharmSUG 2005 Paper FC01 - Implementation of the CDISC SDTM at the Duke Clinical Research Institute, Jack Shostak, Duke Clinical Research Institute (DCRI), Durham, NC

- (17) PharmaSUG 2005 Paper FC03 - Strategies for Implementing SDTM and ADaM Standards, Susan J. Kenny, Maximum Likelihood Solutions, Inc and Octagon Research Solutions, Inc., Chapel Hill, NC
- (18) Michael A. Litzsinger, SCHWARZ BIOSCIENCES, Inc., Research Triangle Park, NC

これらの参考文献の内、

FDA 関連資料は、<http://www.fda.gov/>より、

CDSIC 関連資料は、<http://www.cdisc.org//index.html> より、

また、PharmaSUG 2005 関連資料は、<http://www.pharmasug.org/>

の各 URL より入手できる。また、CDISC 関連資料は CDISC の日本グループが翻訳作業を進めており、日本版が入手可能である。

• 謝辞

Wayne Kubick - Lincoln Technologies

Gary Walker – Quintiles

本報告は、Quintiles の Kansas および RTP (North Carolina)の CDISC に関する Task Force(Gary Walker がリーダー) チームにより作成された。(文責：松岡 淨(Quintiles Transnational Japan K.K.))