

解析業務プロセスにおいて効率的な 仕様書作成とDefine.xmlへの変換

北原 孝志, 東島 正堅

株式会社ACRONET 生物統計部

北西 由武, 吉田 祐樹

塩野義製薬株式会社 解析センター

The efficient preparation of Specification and
its conversion into Define.xml on the process
of statistical analysis

Takashi Kitahara, Masataka Higashijima

Biostatistics Department ACRONET Corporation

Yoshitake Kitanishi, Yuki Yoshida

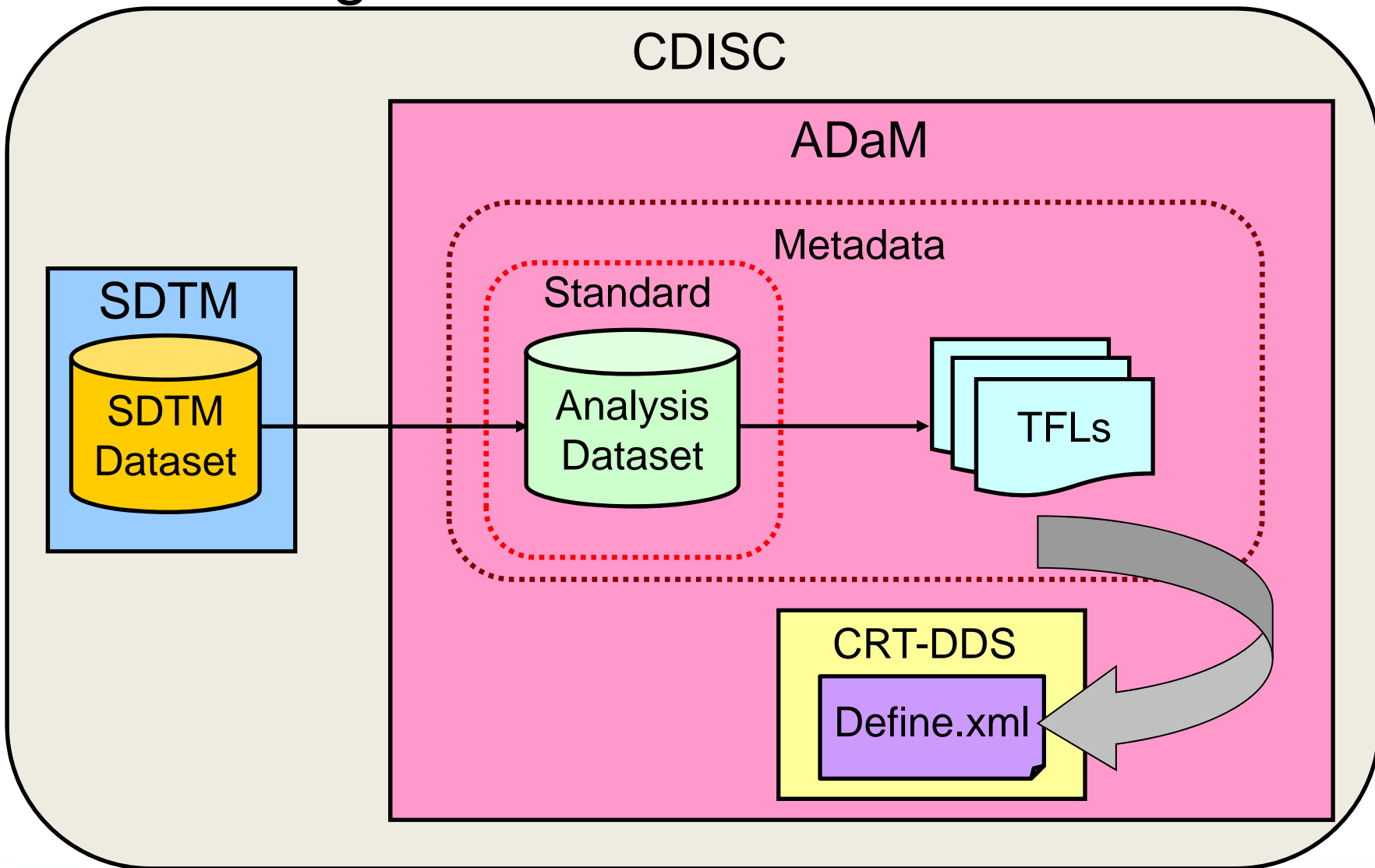
Biostatistics Dept. SHIONOGI & CO., LTD.

要旨：

解析業務プロセスに効果的な仕様書として解析用データセットおよび解析結果のDefine.xmlを作成した。その事例を紹介する。また、Define.xmlの効果的な利用方法を提案する。

キーワード：CDISC, FDA, ADaM, メタデータ, Define.xml

Overview Figure



CDISCとは？

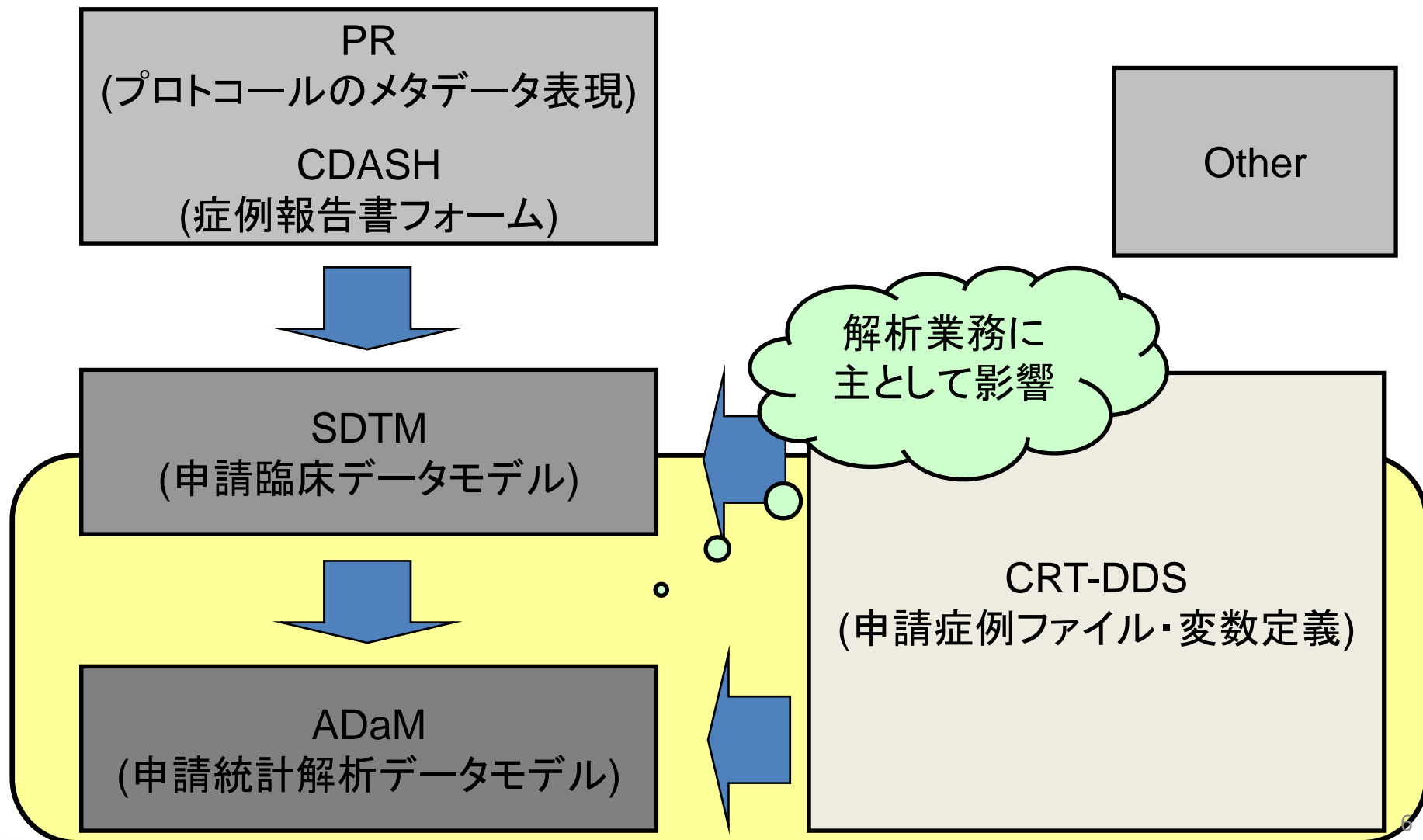
- CDISC is a global, open, multidisciplinary, non-profit organization that has established standards to support the **acquisition, exchange, submission and archive** of clinical research data and metadata.
- The CDISC mission is **to develop and support global, platform-independent data standards** that enable information system interoperability to improve medical research and related areas of healthcare.

<http://www.cdisc.org/>

CDISC標準を導入することの利点

- 規制当局(FDA)
 - 申請資料のより迅速かつ正確なレビューが可能になる
 - 標準化されていないデータや文書では、内容を読み解くのに時間を要してしまう
- 企業
 - データ形式が統一されることによる業務効率化
 - プロジェクト間やパートナー(CRO等)間でのコミュニケーションが容易になる
 - データの統合が容易に出来る

CDISC 各標準のフロー



ADaMとは？

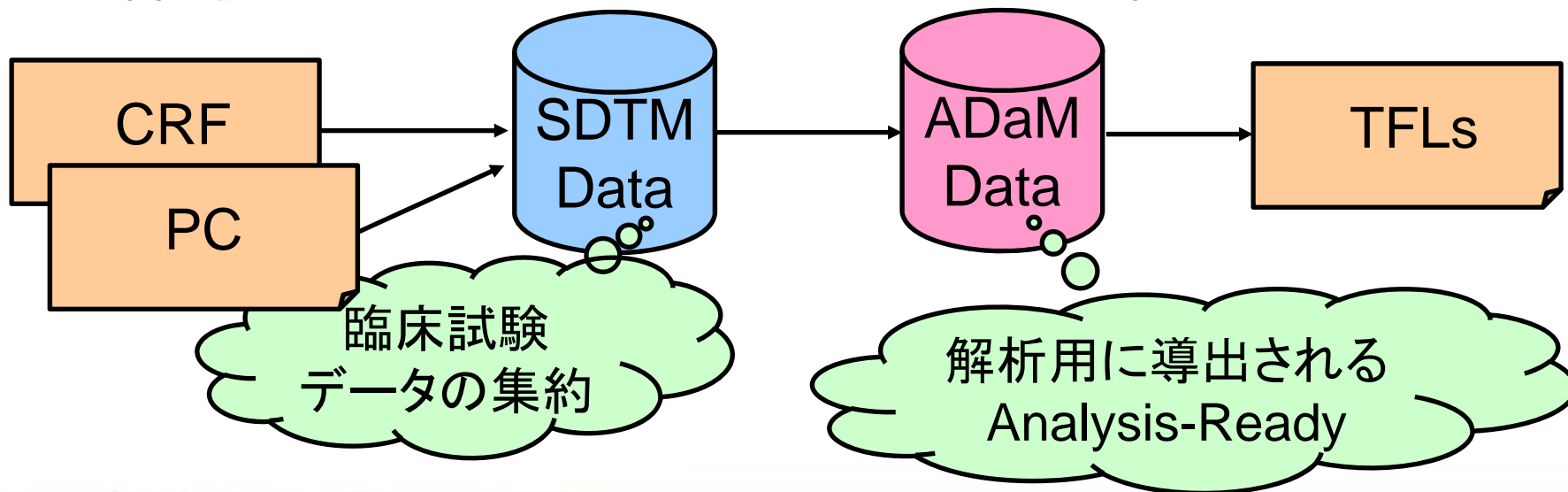
- FDA に医薬品の承認申請を行う際に提出する解析用データセットやメタデータの標準を定めたモデル
 - 統計レビューアがSDTMデータから解析結果を導出するまでのデータの一連の流れを明確に理解できることを目的とする
 - 解析結果を容易に作成するための解析用データセット構造を標準化
 - 解析用データセットを作成するための基本的原則
 - 解析用データセットの構造や命名ルール
 - 変数の命名ルール
 - メタデータの標準化
 - 解析用データセットに対するもの
 - 解析結果に対するもの

ADaMとSDTMの相違は？

・ SDTMとは？

- ・ 臨床試験の内容, 臨床試験で得られたデータをレビューアが明確に理解できることを目的とする
- ・ CRFや日誌で収集された被験者データの標準モデル
- ・ 試験デザイン等のプロトコル情報もデータとして含む

<臨床試験におけるADaMおよびSDTMデータの位置付け>



メタデータとは？

- 一般的な表現
 - データについてのデータ
 - あるデータそのものでなく、そのデータに関連する情報のこと
 - SASデータセットでは、ファイルの更新日時、データセットラベル、変数の長さ、変数タイプ等
 - データの作成日時や作成者、データ形式、タイトル、注釈など

ADaMにおけるメタデータの内容は？

- ADaMにおけるメタデータとは？

解析の文書, 解析結果, 解析に使用されるデータ, およびSDTM-ADaM間の詳細と関連性を記述した仕様書を提供することによって, 情報交換を容易にするもの

データセット
メタデータ

データセット

変数
メタデータ

レコード
メタデータ

TFLs

解析結果
メタデータ

ADaMにおけるメタデータの内容は？

- データセットメタデータ
 - データセット名, データセットの内容, データセットの構造, データセットの分類等
- 変数メタデータ
 - データセット名, 変数名, 変数ラベル, 変数タイプ, 変数フォーマット, 導出ルール等
- レコードメタデータ
 - 例えば, LB(臨床検査値)データにおいて, どの検査項目での処理かを明確にするためのメタデータ
- 解析結果メタデータ
 - 図表番号, タイトル名, 解析に用いた変数名, プログラム命令文等

Define.xmlとは？

- メタデータを集約したXML形式のファイル
- XML形式のファイルとは？
 - Extensible Markup Language
 - 様々なシステム環境で読み取り可能なファイル形式
 - XMLファイル単体を人が読み取るのは難しい
 ⇒ スタイルシートを用いて人が見やすい形式、
 情報を取り出しやすい形式で表示

```

<!-- ***** -->
<!-- Defining ItemGroupDef for data set ADSL -->
<!-- ***** -->
<ItemGroupDef OID="TABLE001"
  Name="ADSL"
  Repeating="Yes"
  IsReferenceData="No"
  SASDatasetName="ADSL"
  Document="SAP 1.1"
  crt:Label="Data for the Demographic Analyses Dataset"
  crt:Structure="one record per subject"
  crt:DomainKeys="STUDYID, USUBJID"
  crt:Class="ADSL"
  crt:DClass="ADaM"
  crt:ArchiveLocationID="Location.TABLE001"
>
    
```

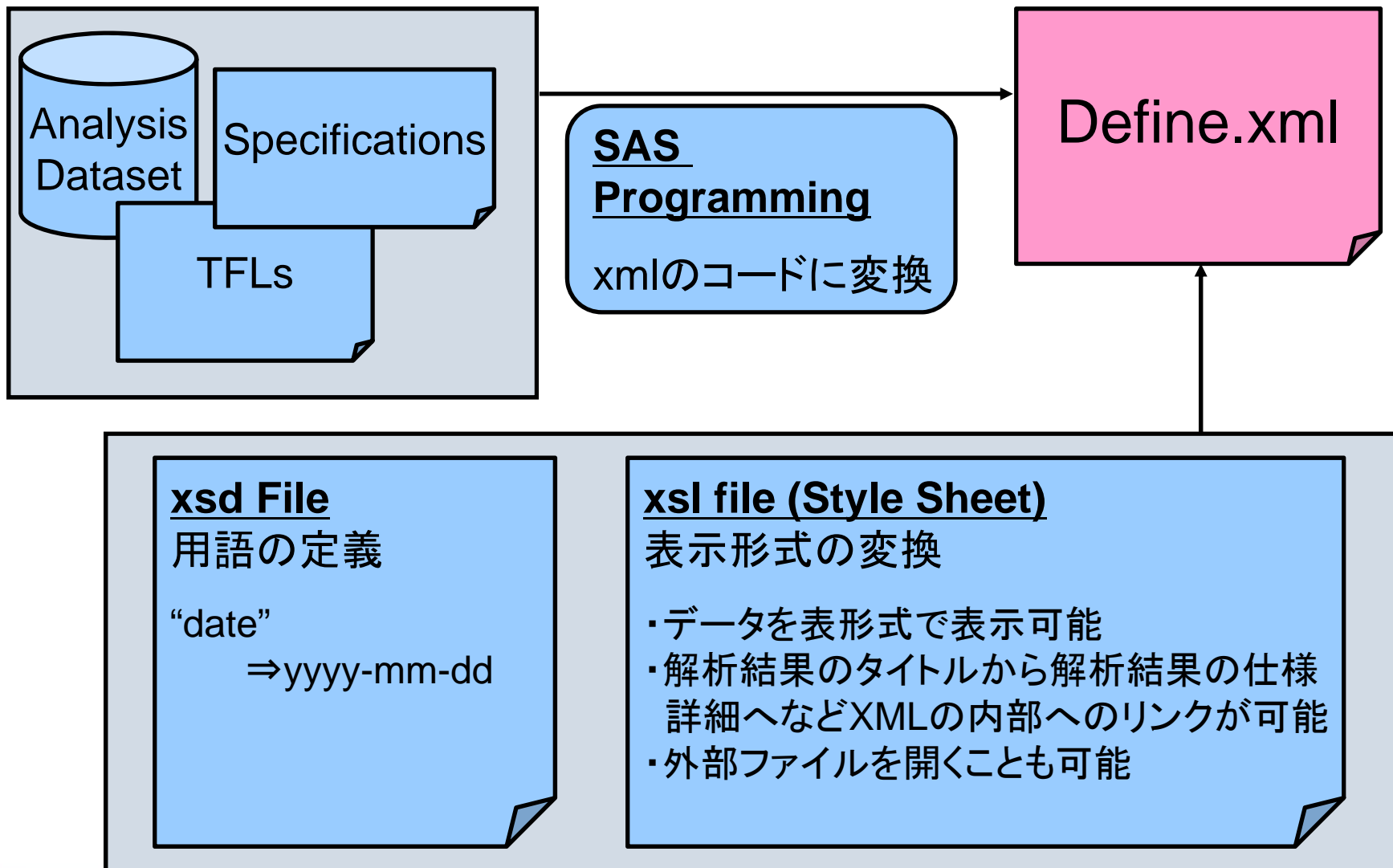
スタイルシートを適用すると...

Analysis Datasets for SAS-001						
Dataset	Dataset Description	Dataset Location	Dataset Structure	Key Variables of Dataset	Class of Dataset	Documentation
ADSL	Data for the Demographic Analyses Dataset	adsl.xpt	one record per subject	STUDYID, USUBJID	ADSL	SAP 1.1
ADEF	Contains the data for the Primary End-Point	adef.xpt	one record per subject per VISIT	STUDYID, USUBJID	BDS	SAP 2.2
ADAE	Contains the data for the Adverse Event Analyses	adae.xpt	one record per subject per each AE recorded in SDTM AE domain	STUDYID, USUBJID, AESEQ	ADAE	SAP 3.3

CRT-DDSとDefine.xml

- Define.xmlはCRT-DDSで定義され、FDA申請の際に提出する解析用データセットの仕様等を電子上で閲覧できるファイル
 - 解析用データセットの内容(変数, 構造, 内容等)や解析結果の内容を明確に提示し、規制当局側のレビューアの手助けをする

Define.xml作成までのフロー



CDISCパイロットスタディの概要

- 2006～2007年度にFDAと製薬企業が共同で実施
- CDISCが提示するモデル(SDTMおよびADaM)が、製薬企業側とFDAのレビューアの必要とするところに、どの程度マッチするかを確認することが目的
- CDISC会員のみ閲覧可能
- ADaMのドキュメントVersion2.0を参考に作成
(現在: Version 2.1が最新(2009.12.17リリース))

Define.xmlの作成背景

CDISCホームページよりパイロットスタディのDefine.xml作成
SASプログラム, Define.xmlおよび総括レポートをダウンロード
(CDISC会員限定でダウンロード可)



プログラムをそのまま流せば作成できるものだと思っていたが...
⇒最新のADaMメタデータ形式に従っていない
⇒Input情報およびスタイルシートを最新にする必要あり
⇒プログラムを部分的に変えて作成できるという訳ではなかった



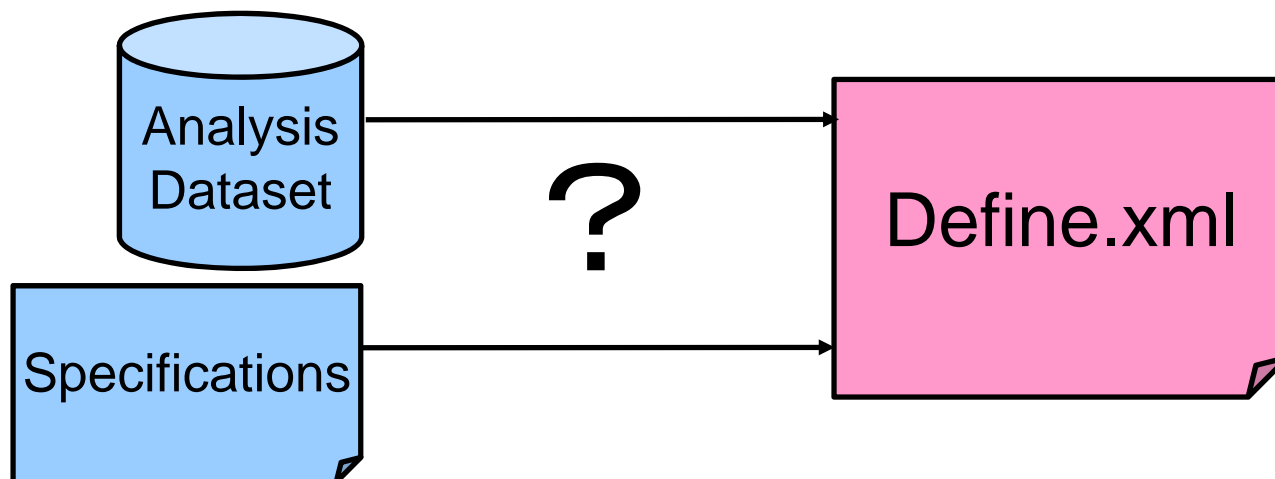
Define.xml作成プログラムの再構築を決意！

作成手順の検討

- パイロットスタディSASプログラムの内容①
 - Spreadsheet (Excelファイル)を読み込んで作成
 - ⇒ 解析用データセットおよび解析結果の仕様書を Define.xmlに出力される形にして作成する事とした
 - ADaMドキュメントの最新バージョンに合わせる必要性があった
 - 既存の仕様書から情報を追加する事で作成できると判断した

作成手順の検討

- パイロットスタディSASプログラムの内容②
 - Spreadsheetから作成する以外に解析用データセットから作成するSASプログラムも存在



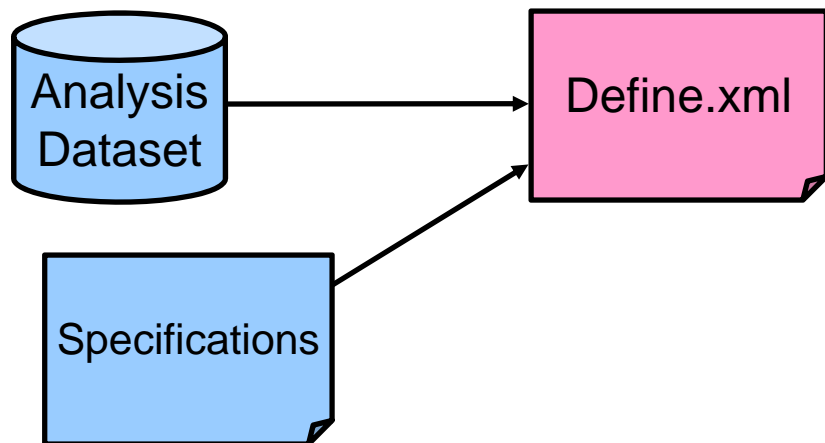
作成手順の検討

① ②のどちらの方法が最適かの検討を行った

作成方法	①仕様書から作成する	②解析用データセットから作成する
Pros	変換処理の詳細等, データの情報が集約されている	データセットそのものの情報を得る事により, Define.xmlに出力する内容との不整合がない
Cons	解析用データセットと Define.xmlとの内容に不整合がおきる可能性がある	変換処理の詳細等の情報を後から追加する必要がある

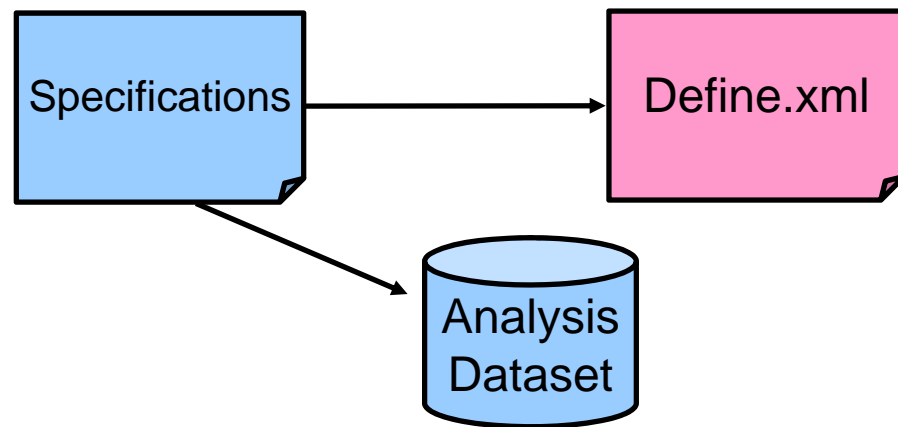
作成手順の検討(Consを埋めるためには)

・解析用データセットから作成



変換処理の詳細等の情報を後から追加する必要がある
 ⇒ 結局仕様書から情報を得なければならない

・仕様書から作成



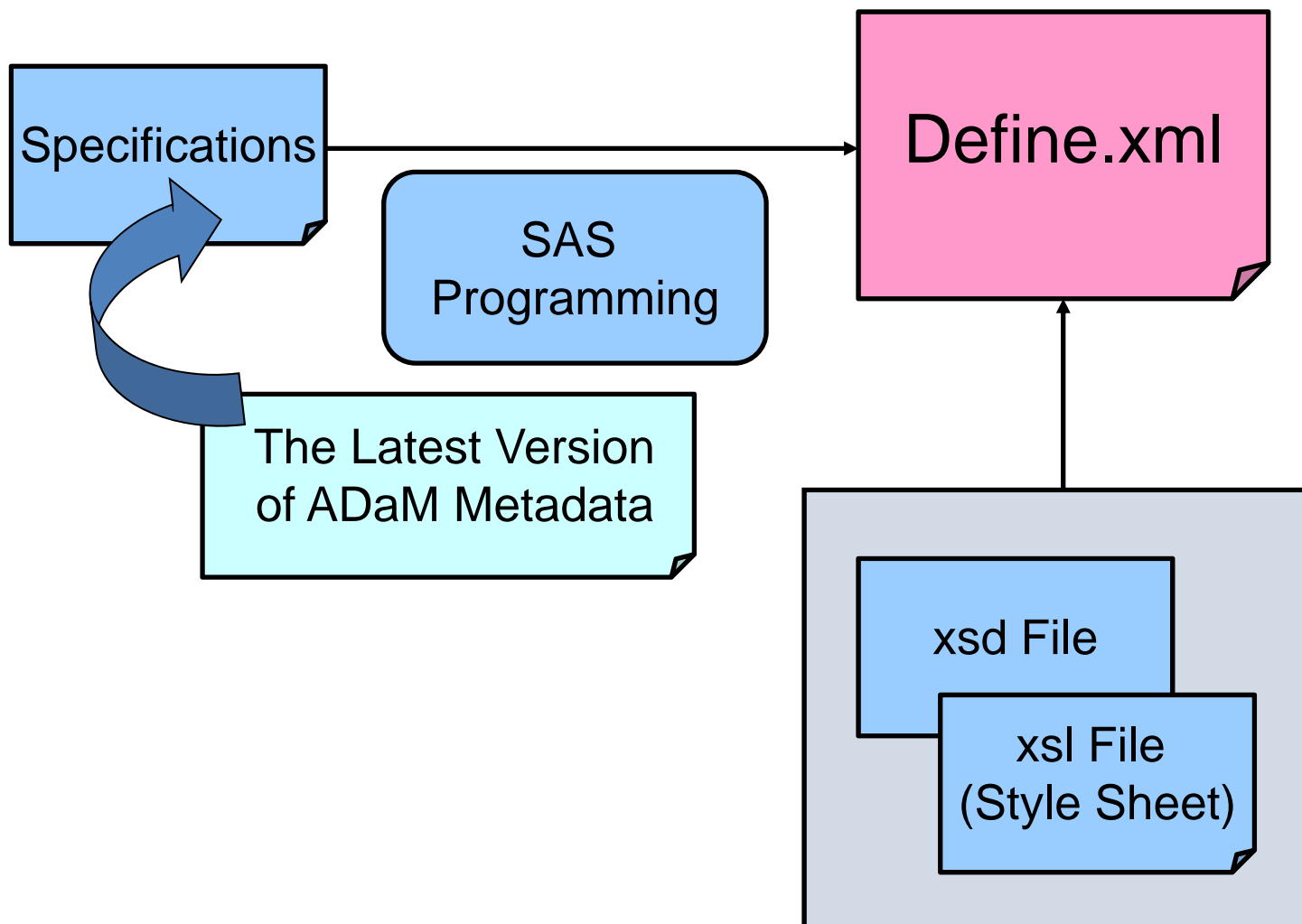
解析用データセットとDefine.xmlとの内容に不整合がおきる可能性がある
 ⇒ 仕様書から機械的にデータセットに情報を渡す事で不整合がなくなる

仕様書から作成することが効率的と判断した

仕様書の作成

- ADaMで定められた最新バージョンのメタデータ形式に合わせて作成した
- 既存の仕様書から大きく変更する事なく作成
- 解析用データセット仕様書
 - データセットに関連するADaMメタデータ形式
 - ① データセットメタデータ
 - ② 変数メタデータ
 - ③ レコードメタデータ
- 解析結果作成仕様書
 - 解析結果に関連するADaMメタデータ形式
 - ① 解析結果メタデータ

作成フローまとめ



例示

Analysis Results (Summary) for SAS-001

[Table 1-1](#) - Demographics and Baseline Characteristics

[Table 2-2](#) - Analysis of Change in Primary End-Point

[Table 3-3](#) - Summary of Treatment Emergent Adverse Events

Go to the top of the [define.xml](#)

Date of document generation (2011-06-29T19:47:31)

解析結果の
Define.xml表示

Display Identifier	Table 2-2
DisplayName	Analysis of Change in Primary End-Point
Result Identifier	Analysis of Change in Primary End-Point
Parameter	Total Score
Parameter Code	TS
Analysis Variable	CHG
Reason	Primary efficacy analysis as pre-specified in protocol
Dataset	Contains the data for the Primary End-Point (ADEF)
Selection Criteria	FASFL='Y'
Documentation	SAP 2.2 , testprotocol TFL T2.2 , Calculate p-value using two sample t-test.
Programming Statements	PROC TTEST; CLASS TRIAN; VAR CHG; RUN ;

Go to the [Analysis Results \(Summary\)](#)

問題点/解決策

- ① 全ての解析結果が1つのpdfファイルに集約されることを想定したとき、見たい箇所をすぐに表示するのが困難である

⇒解析結果のpdfファイルにハイパーリンクをつける際、開いたpdfファイルへの文字列検索機能をつけることで、(例. 図表タイトル)大容量の解析結果の見たい場所をすぐ表示できる

Display Identifier	Table 3-3
DisplayName	Summary of Treatment Emergent Adverse Events
Result Identifier	Summary of Treatment Emergent Adverse Events
Parameter	-
Parameter Code	-
Analysis Variable	AETERM
Reason	Safety analysis as pre-specified in protocol
Dataset	Contains the data for the Adverse Event Analyses (ADAE)

クリック



検索する語句:
現在の文書における 3-3

結果:
2 文書で 2 件を検索しました

結果:

- AnalysisResults.pdf
- Table 3-3 Summary of Treatment Emergent
- AnalysisResults.pdf

Table 3-3 Summary of Treatment Emergent Adverse Events Safety population		
	Active N=99	Placebo N=99
Adverse Events		
Number of Patients	99	99
(Number of Events)	(999)	(999)
Incidence (%)	99.9	99.9
95% Confidence Interval (%)	(99.9, 99.9)	(99.9, 99.9)

検索したページを表示

問題点/解決策

- ② Define.xmlの作成手順やファイルの構成等が複雑である

⇒ Define.xml作成者へのトレーニング資料を作成

トレーニング資料の内容

メタデータの概要

Define.xmlの構成, 作成手順等

問題点/解決策

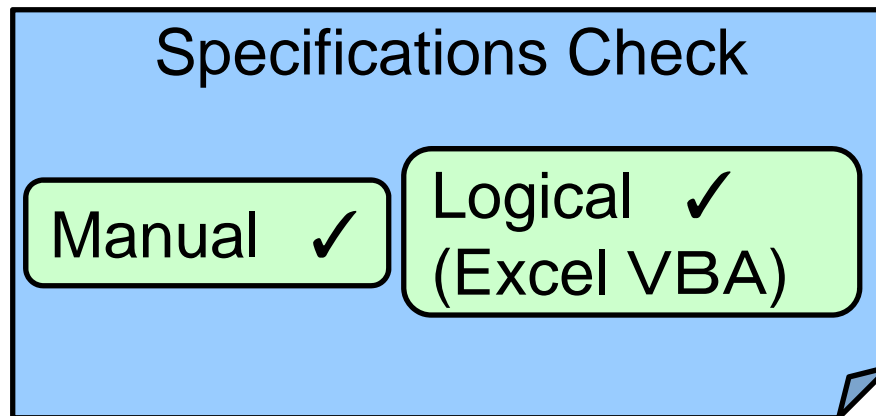
③ xmlに関する知識を持った担当者が少ない

⇒プログラムのメンテナンスや問題点を早期に対応できるように、体制を整えておく(Define.xmlの管理者を明確にしておく)

問題点/解決策

- ④ 仕様書のManual/Logicalチェックの切り分けが必要であった

⇒標準の仕様書を再考するのに良い機会であったので、Manual/Logicalチェックの切り分けを行い、Logicalチェックを行える箇所は、ExcelのVBAを付与し、業務の効率化を図った



提案

- Define.xmlを申請資料としてだけでなく、社内での運用ツールとして用いてはどうか
 - 1つのファイルに全ての情報が標準化され集約されることにより、社内レビューの時間短縮が見込まれる
 - 過去の試験の解析用データセット・解析結果の仕様、ならびに解析結果が容易に閲覧可能となる
 - CDISCに準拠したデータモデルの提出が規制当局から義務付けられる可能性があるので、検討しておくべきである

結論

- Define.xmlを作成する際に、仕様書から作成することを提案した
 - Define.xmlを作成した今回の検討によって、社内標準の仕様書を見直す良い機会であった
- Define.xml作成には、xmlに精通したプログラム開発者が必要となる
 - 育成 or 外部リソースの活用
- Define.xmlを申請資料としてだけでなく、解析業務の中で有効利用する

略語一覽

ADaM	<u>A</u>nalysis <u>D</u>ata <u>M</u>odel
CDASH	<u>C</u>linical <u>D</u>ata <u>A</u>cquisition <u>S</u>tandards <u>H</u>armonization
CDISC	<u>C</u>linical <u>D</u>ata <u>I</u>nterchange <u>S</u>tandards <u>C</u>onsortium
CRT-DDS	<u>C</u>ase <u>R</u>eport <u>T</u>abulation and <u>D</u>ata <u>D</u>efinition <u>S</u>pecification
FDA	<u>F</u>ood and <u>D</u>rug <u>A</u>dministration
PR	<u>P</u>rotocol <u>R</u>epresentation
SDTM	<u>S</u>tudy <u>D</u>ata <u>T</u>abulation <u>M</u>odel