

解析しやすい医薬データにするためのSASによるデータ加工

# 標準化を目指した医薬データの定義

CDISC標準データモデルの理解をもとに

中外臨床研究センター  
バイオメトリクス部  
松丸一郎

- 1 . 医薬データの現状と標準化の重要性
- 2 . CDISCとは？
- 3 . CDISCのデータ定義を理解する  
(SDTMを中心に)
- 4 . 医薬データの標準化を目指して…

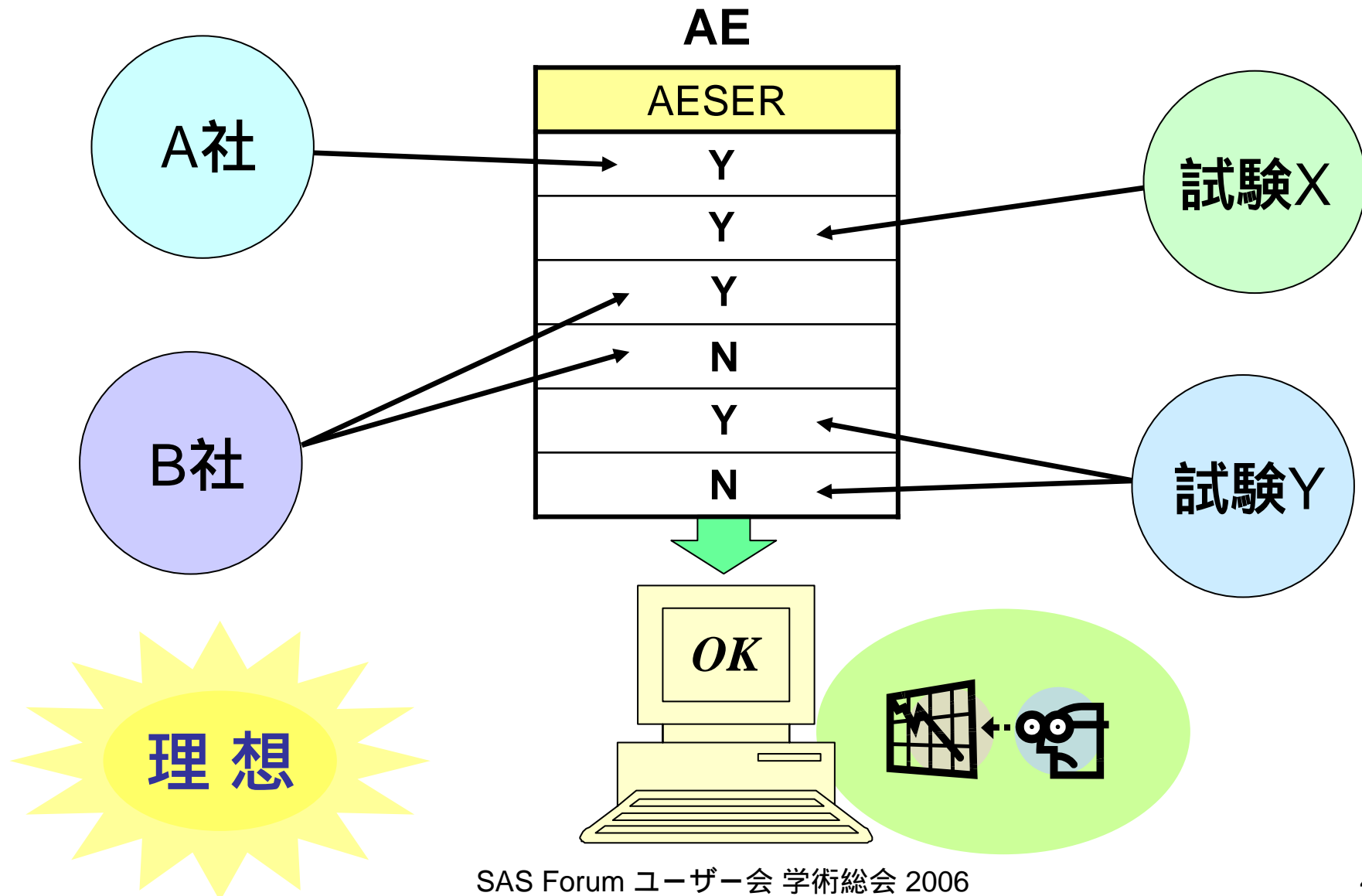
**1 . 医薬データの現状と標準化の重要性**

**2 . CDISCとは？**

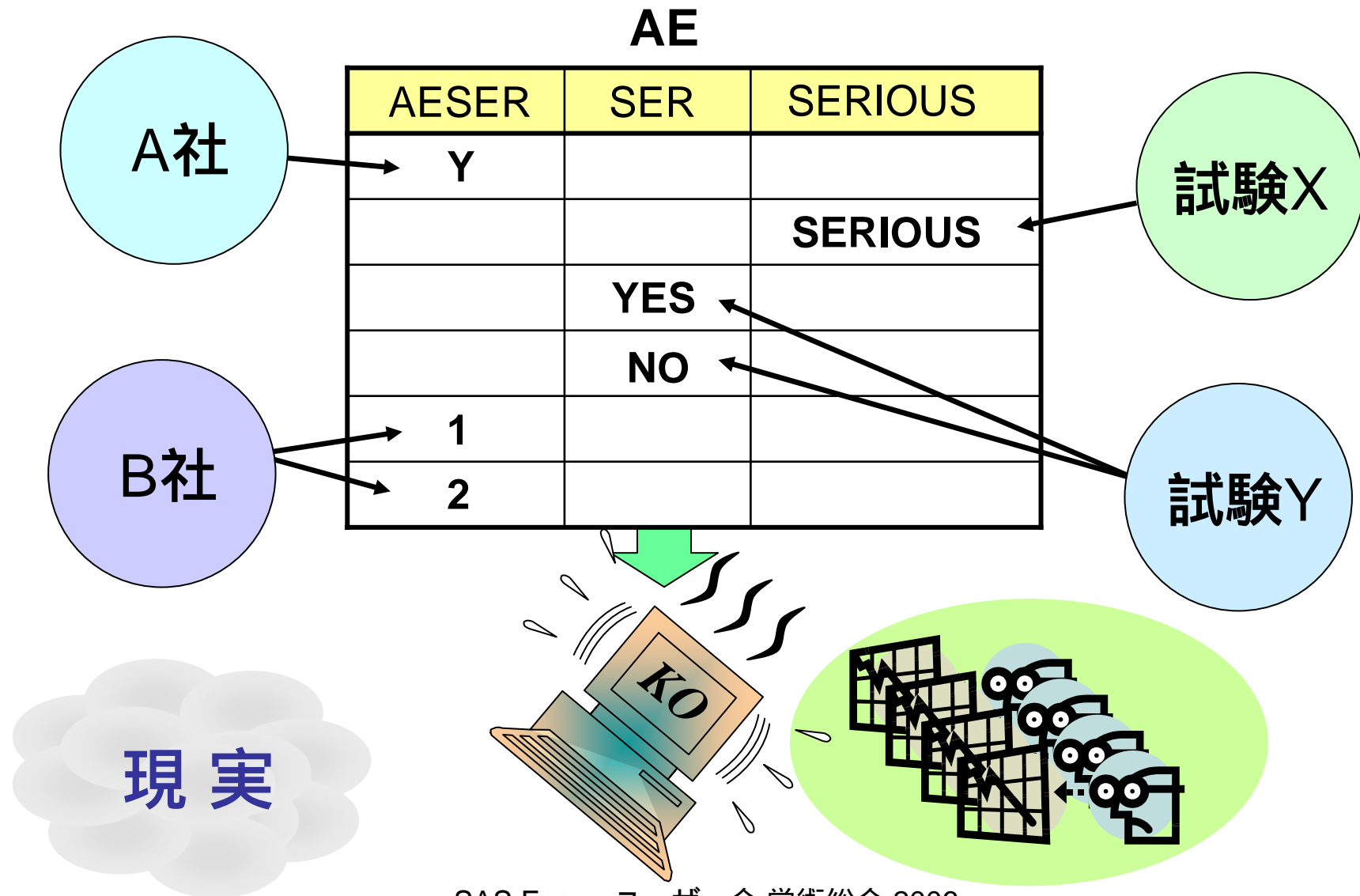
**3 . CDISCのデータ定義を理解する  
(SDTMを中心に)**

**4 . 医薬データの標準化を目指して…**

# グローバル開発やメタ解析を実施してみると・・・



# グローバル開発やメタ解析を実施してみると・・・



# 各社各様の用語・定義・値・コード・ラベル・フォーマット

## Medical History の例

### 用語・ラベル

病歴	Clinical History
既往歴	Past Medical History
既往症	Previous Disease, History of Illness

### 定義

手術歴を含む  
手術歴を含まない( 前治療)

現病歴, 既往歴, 家族歴, ...

### 項目名・値・コード・ラベル・フォーマット

既往歴がありますか？

はい	1: はい	1: はい
いいえ	2: いいえ	0: いいえ

Any Medical History?

YES	NO	Y
NO	YES	N

既往歴(の有無)

あり	なし	有
なし	あり	無

Medical History

PRESENT	EXIST
ABSENT	NOT EXIST

### 構造

合併症から独立したデータセット(別データセット名)

同じ変数名  
異なる変数名

合併症と同じデータセット

既往症・合併症識別のための変数名・値は様々

# Case 1: 同じ変数名, 異なる属性の変数を含む 縦方向の結合 (SET)

**A**

*Numeric*

STUDYID	DOMAIN	USUBJID	AESEQ	AETERM	AETERM_J
STUDY01	AE	STUDY01-02-01	1	PYREXIA	発熱
STUDY01	AE	STUDY01-02-01	2	ALOPECIA	脱毛症
STUDY01	AE	STUDY01-02-02	1	DIARRHOEA	下痢
STUDY01	AE	STUDY01-03-01	1	SYNCOPE	失神

DATA AE ;  
SET A B ;  
RUN ;

**B**

*Character*

STUDYID	DOMAIN	USUBJID	AESEQ	AETERM	AETERM_J
STUDY02	AE	STUDY02-03-01	"1"	DYSURIA	排尿困難
STUDY02	AE	STUDY02-03-01	"2"	TINNITUS	耳鳴
STUDY02	AE	STUDY02-03-01	"3"	DYSPHAGIA	嚥下障害

**ERROR:** 変数 AESEQ は  
文字と数値の両方に  
定義されています。

NOTE: エラーが発生した  
ため、このステップの  
処理を中止しました。

WARNING: データセット  
WORK.AE は不完全  
です。処理終了時の  
データセットは 0 オブザ  
ベーション、6 変数です。

*AESEQのタイプが異なっていた !  
(数値型・文字型)*

```
AESEQ_N = INPUT(AESEQ, BEST.); DROP AESEQ ;  
RENAME AESEQ_N = AESEQ ;
```

## Case 2: 識別変数の一部が欠如したまま 横方向の結合 (MERGE)

A社のデータセット X1, Y1

BY変数:

STUDYID, USUBJID, SEQ\_J

B社からA社に転送されたデータセット X2, Y2

BY変数:

STUDYID, USUBJID, PAGE, SEQ\_J

B社からのデータセットX2, Y2を、いつものX1, Y1と同じつもりでマージ

```
DATA Z2 ;  
  MERGE X2 Y2 ;  
× BY STUDYID USUBJID SEQ_J ;  
RUN;
```

**BY STUDYID, USUBJID, PAGE, SEQ\_J ;**

“ERROR(なし), WARNING(なし)”

多対多のマージが生じた場合

**NOTE:** MERGE ステートメントに  
BY 値を繰り返すデータセットが  
複数あります。

1対1 または 1対多の  
マージだった場合

(オブザベーション数怪しい?)  
(何かありました?)

# Case 3: 誤った識別変数による 横方向の結合 (MERGE)

## CM (CONCOMITANT MEDICATIONS)

STUDYID	DOMAIN	USUBJID	PAGE_J	SEQ_J	CMTRT_J	CMINDC_J	AEPAGE_J	AESEQ_J
STUDY01	CM	STUDY01-02-01	51	1	DICLOFENAC	ADVERSE EVENT	101	1
STUDY01	CM	STUDY01-02-01	51	2	LOPERAMIDE	ADVERSE EVENT	102	1

BYステートメントで  
指定する変数に注意！！

## AE (ADVERSE EVENTS)

STUDYID	DOMAIN	USUBJID	PAGE_J	SEQ_J	AETERM	AETERM_J
STUDY01	AE	STUDY01-02-01	101	1	PYREXIA	発熱
STUDY01	AE	STUDY01-02-01	101	2	ALOPECIA	脱毛症
STUDY01	AE	STUDY01-02-01	102	1	DIARRHOEA	下痢
STUDY01	AE	STUDY01-09-05	102	2	SYNCOPE	失神

RENAME PAGE\_J = AEPAGE\_J SEQ\_J = AESEQ\_J ;

ERROR(なし)  
WARNING(なし)  
NOTE:  
WORK.CASE2A\_CM  
から 2 オブザベーション  
  
WORK.CASE2B\_AE  
から 4 オブザベーション  
  
WORK.CM\_IND\_AE  
は 6 オブザベーション、  
11 変数です。

# 不具合の検出には限界がある

SAS[ログ]ウィンドウ中の

「ERROR」, 「WARNING」, 「繰」,  
「初期」, 「0 オブザベーション」

等の文字列検索で、検出できる場合もあるが、  
ERRORやWARNINGを返さない場合もあり、  
限界がある。

# DAY計算

例：日本の試験と海外の試験で  
DAY計算が合わない？

日本の試験：起点 Day 0\*      Day 7: 7日後

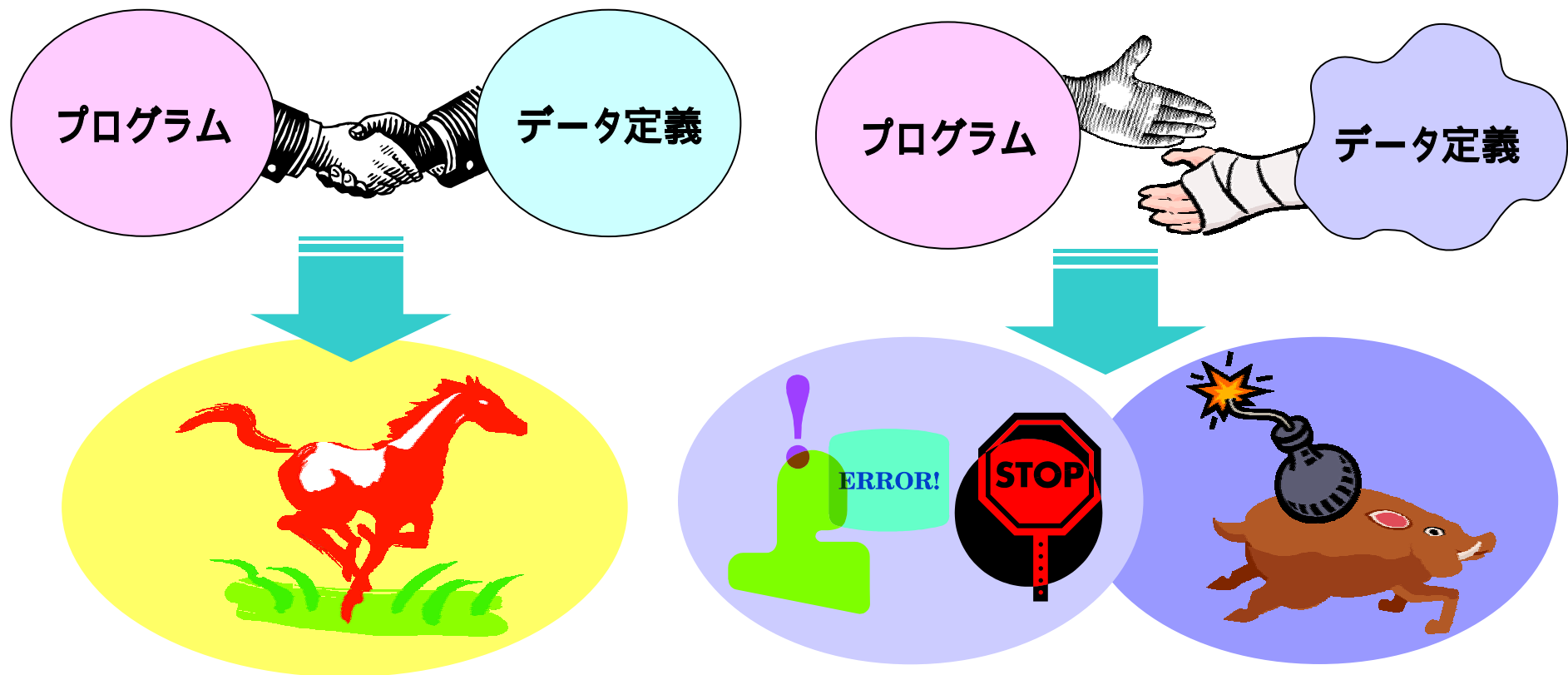
海外の試験：起点 Day 1      Day 7: 7日目(6日後)

(前日はどちらも Day -1)

\* あくまで1つの例。すべての日本の試験がこれで計算しているわけではない。  
日本の多くの試験でも上記の海外試験と同じDAY計算が用いられている。

# データ定義の標準化の重要性

プログラムがきちんと動いて、正しい値が返るのも、  
“想定内”のデータであればこそ…



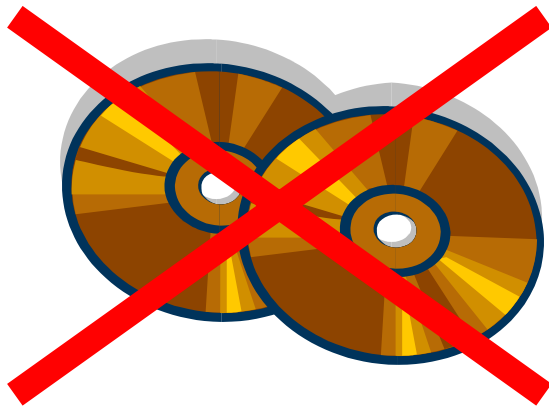
1 . 医薬データの現状と標準化の重要性

2 . CDISCとは？

3 . CDISCのデータ定義を理解する  
(SDTMを中心に)

4 . 医薬データの標準化を目指して…

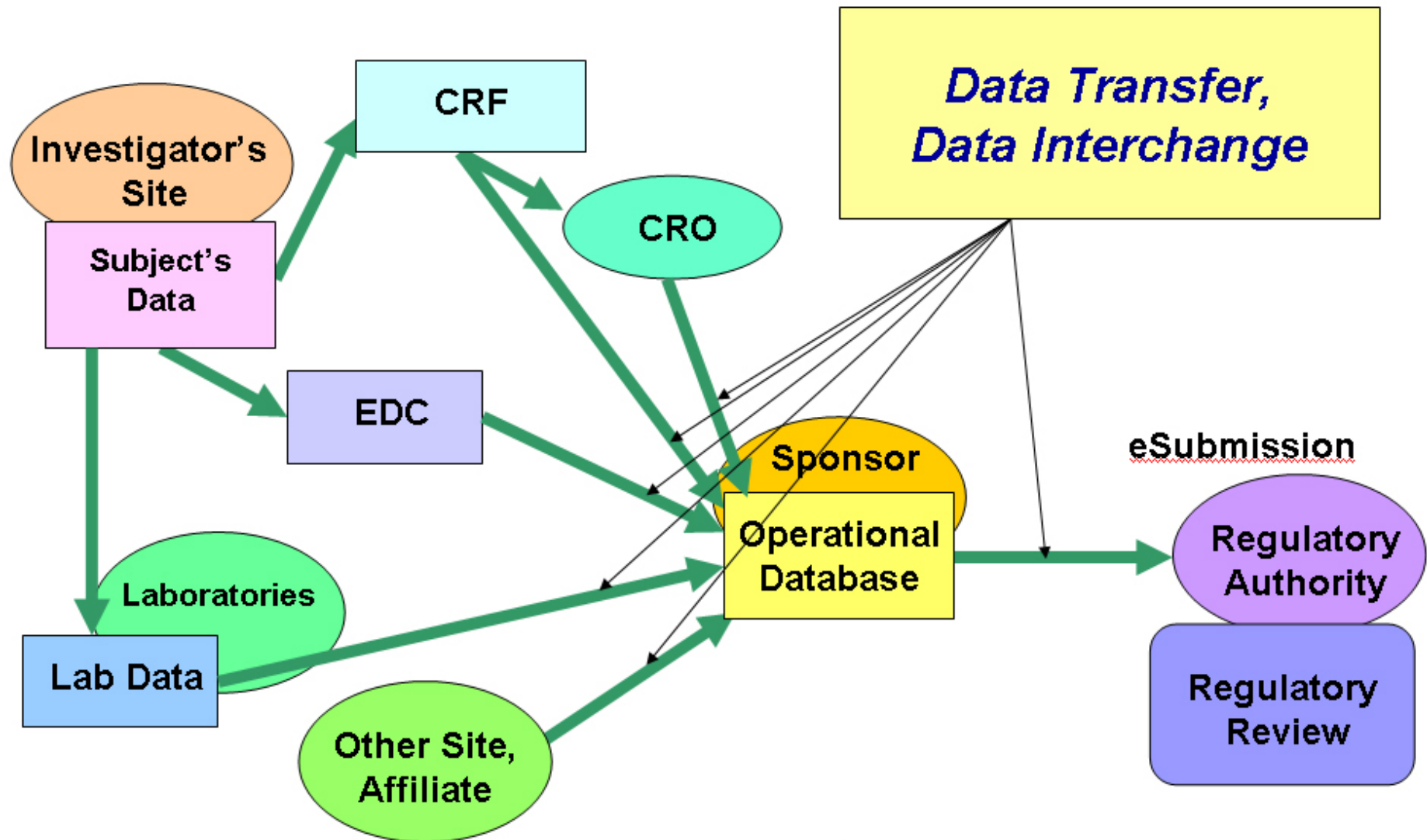
# What s CDISC ?



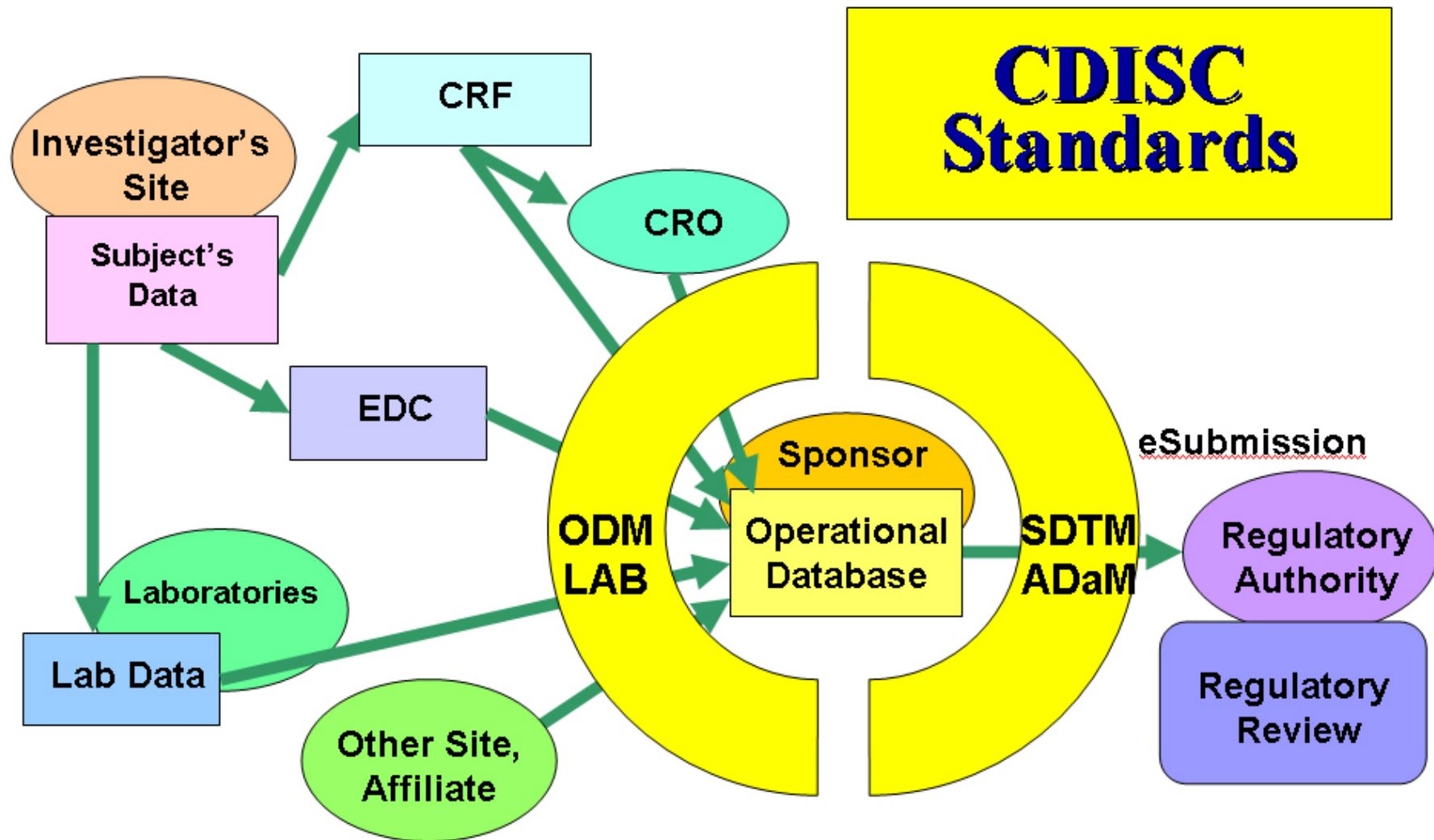
*“New Type of Disc?”*  
**No !**

***Clinical Data Interchange Standards Consortium***

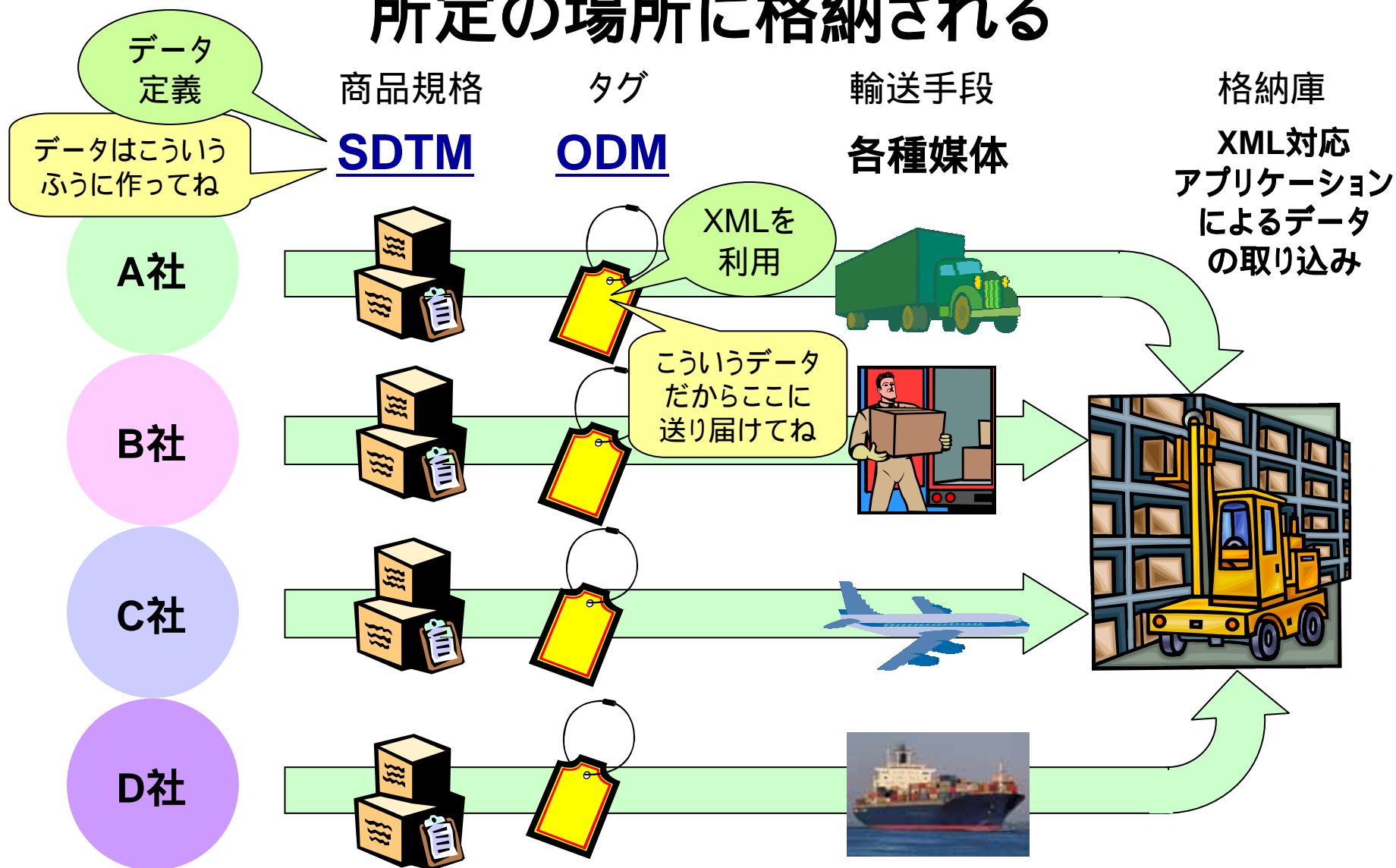
# Clinical Study Data Flow



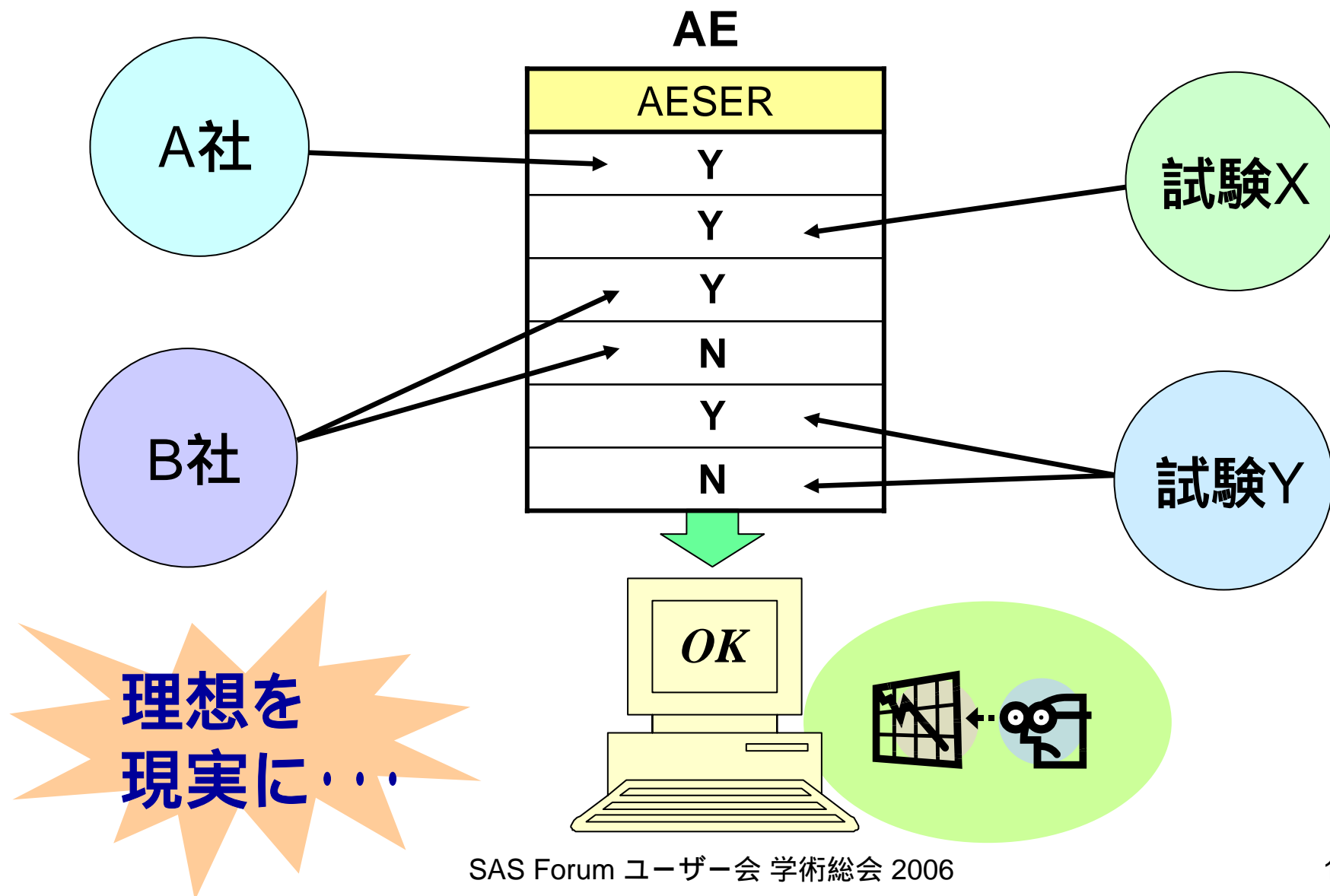
# Clinical Study Data Flow and CDISC Standards



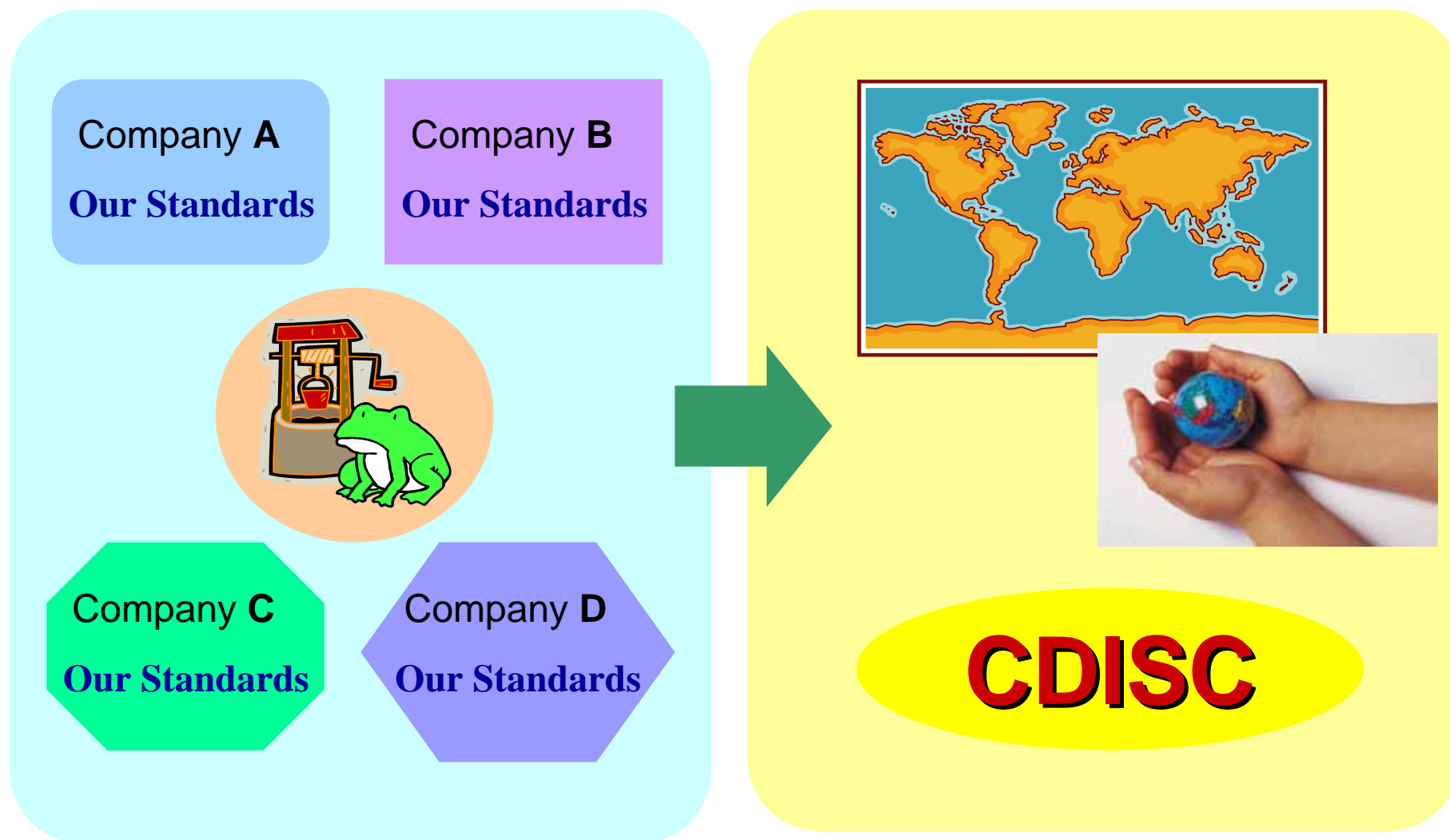
# 同じ規格の商品に同じタグを付けて送れば 所定の場所に格納される



# Global Standards will bring Efficiency to us



# Global Standards, As Gold Standards



1 . 医薬データの現状と標準化の重要性

2 . CDISCとは？

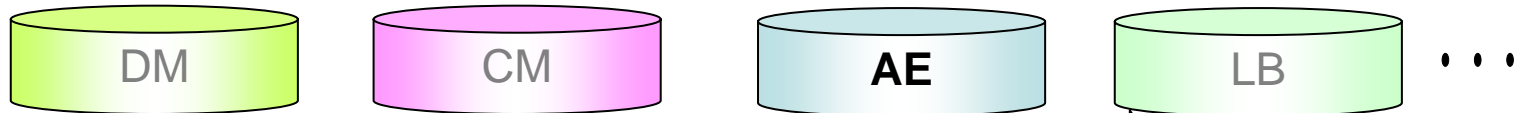
3 . CDISCのデータ定義を理解する  
(SDTMを中心に) (+ ADaM)

4 . 医薬データの標準化を目指して…

**SDTM:** Study Data Tabulation Model  
**ADaM:** Analysis Dataset Model

# To Understand SDTM ...

## Domains



## Variables

STUDYID, DOMAIN, USUBJID, AESEQ, AETERM, AESEV, **AESER**, AEPATT, ...

One record per adverse event per subject

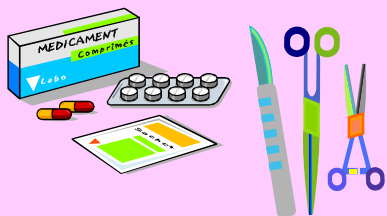
## Definitions

• Variable Name	• AESER	
• Variable Label	• Serious Event	
• Type	• Char	• <b>Char</b> , Num
• Controlled Terms or Format	• **Y, N	• * sponsor-controlled terminology • ** external controlled terminology
• Origin	• CRF or Derived	• CRF, Derived, Sponsor Defined
• Role	• Record Qualifier	• Identifier, Topic, Timing, Synonym Qualifier, Grouping Qualifier, Record Qualifier, Variable Qualifier
• CDISC Notes	• Is this a serious event?	
• Core	• Exp	• <b>Req</b> (uired), <b>Exp</b> (ected), <b>Perm</b> (issible)
• References		

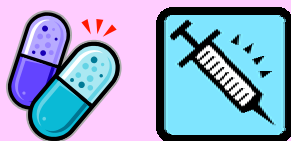
# Domains by Observation Class

## Interventions

**CM:** Concomitant Medication



**EX:** Exposure



**SU:** Substance Use



## Special Purpose

**DM:** Demographics

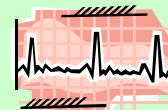


**CO:** Comments



## Findings

**EG:** ECG Test Results



**IE:** Inclusion/Exclusion Exceptions

**Yes/No**

**LB:** Laboratory Test Results



**PE:** Physical Examination

**QS:** Questionnaires



**SC:** Subject Characteristics



**VS:** Vital Signs



## Events

**AE:**  
Adverse Events



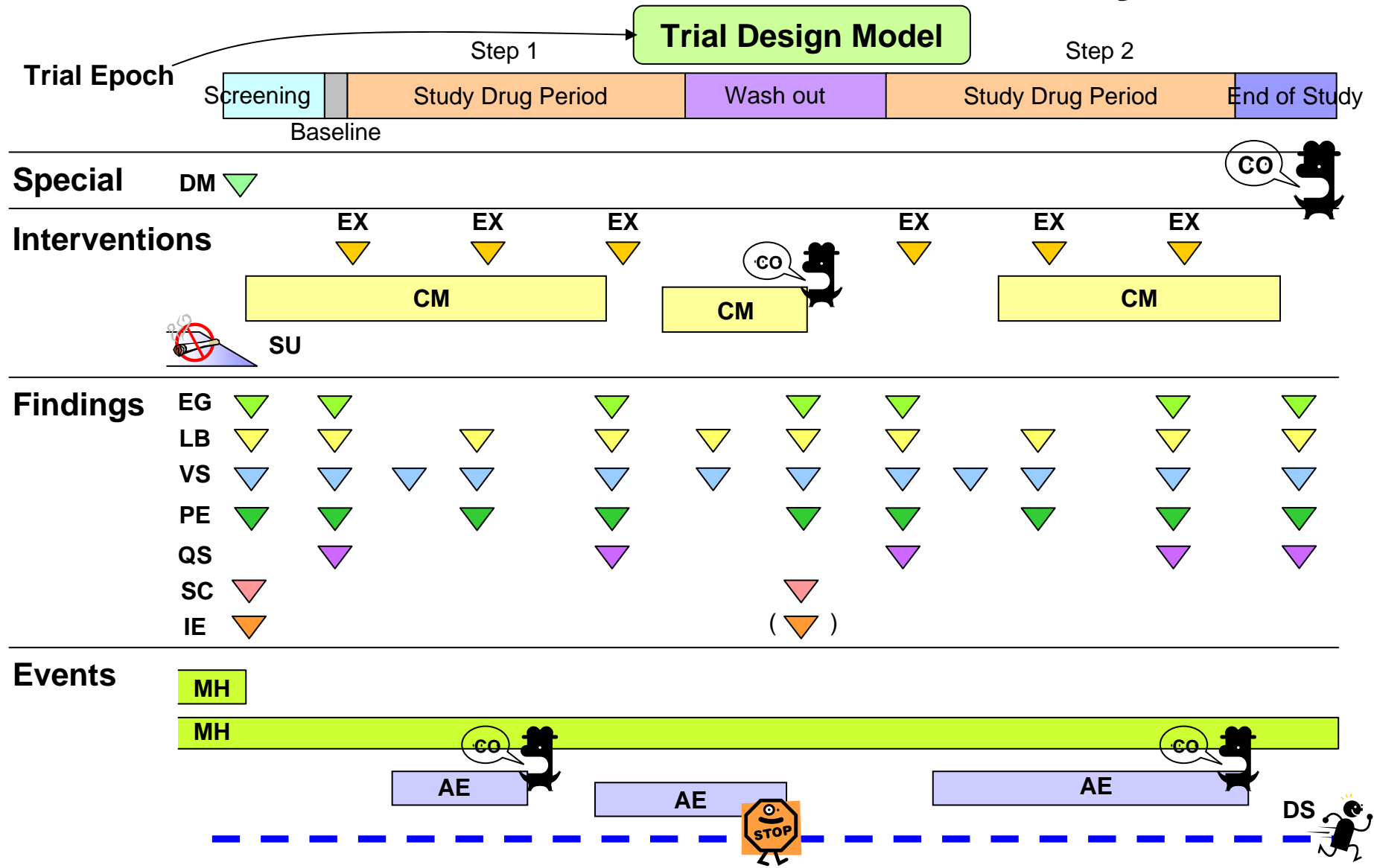
**DS:** Disposition



**MH:**  
Medical History



# Domains in a Clinical Study



# **SDTM** (*Study Data Tabulation Model*)

## **“Implementation Guide”のデータ定義の見方**

Domain: **AE** Variable: **AESER**

Variable Name	Variable Label	Type	Controlled Terms or Format	Origin	Role	CDISC Notes	Core	References
AESER	Serious Event	Char	**Y, N	CRF or Derived	Record Qualifier	Is this a serious event?	Exp	

変数名

変数  
ラベル

文字型？  
数値型？

用語・  
フォーマットは？

どこ由来の  
データ？

どんな役割  
のデータ？

留意点

必須変数？  
必要に応じて使用？

参照  
出典

# SDTMを理解するためのコツ

1. 同じ変数は同じ意味・使い方 (Standardの原則)。
2. 同じ接尾語は同じ使い方。 (いろいろな使い方が認められる場合あり)
3. Structure (レコードの持ち方) が規定されている。 (例外あり)
4. Core は次の3種類
  - “Req(uired)” : 変数・値ともdomainに必須,
  - “Exp(ected)” : 変数はdomainに必須、値はnullもあり得る,
  - “Perm(issible)” : 変数・値とも必要に応じて
5. Type: “Char: 文字型”, ”Num: 数値型” に注意。
6. Controlled Terms / Formatが規定されているもの、治験依頼者が独自に定めてよいものがある。
7. Date/Time , DurationはISO8601 (文字型)。
8. “コメント” は「CO」domainにデータがおかれる。  
(特殊なレコードの持ち方のdomainに注意)

1. 同じ変数は同じ意味・用い方

2. 同じ接尾語の変数は同じ用い方

### Identifiers

Variable Name	Variable Label	Type
<b>STUDYID</b>	Study Identifier	Char
<b>DOMAIN</b>	Domain Abbreviation	Char
<b>USUBJID</b>	Unique Subject Identifier	Char

### Timing Variables

Variable Name	Variable Label	Type
<b>VISITNUM</b>	Visit Number	Num
<b>VISIT</b>	Visit Name	Char
<b>VISITDY</b>	Planned Study Day of Visit	Num

### Identifiers

Variable Name	Variable Label	Type
<b>--SEQ</b>	Sequence Number	Num
<b>--GRPID</b>	Group ID	Char
<b>--SPID</b>	Sponsor ID	Char



## 4. Core

**Req(uired)** :

変数・値ともdomainに必須

**Exp(ected)** :

変数はdomainに必須、値はnullもあり得る

**Perm(issible)**:

変数・値とも必要に応じて

## 5. Type (Char: 文字型 , Num: 数値型)

**EX (Exposure):** 変数名に「DOS」を含む変数 (抜粋)

Variable Name	Variable Label	Type	Origin	CDISC Note	Examples	Core
<b>EXDOSE</b>	<b>Dose per Administration</b>	<b>Num</b>	CRF or Sponsor Defined	Amount of EXTRT administered or given.		<b>Exp</b>
<b>EXDOSTXT</b>	<b>Dose Description</b>	<b>Char</b>	CRF or Derived	Dosing amounts or a range of dosing information collected in text form	<b>「200-400」</b>	<b>Perm</b>
<b>EXDOSU</b>	Dose Units	Char	CRF or Sponsor Defined	Units for EXDOSE and EXDOSTOT.	ng, mg, mg/kg	Exp
<b>EXDOSFRM</b>	Dose Form	Char	CRF or Sponsor Defined	Dose form for EXTRT	TABLET, LOTION,.	Perm
<b>EXDOSFRQ</b>	Dosing Frequency Per Interval	Char	CRF or Sponsor Defined	Usually expressed as the number of dosings given per a specific interval.	BID, QID	Perm

### 3. Structure (レコードの持ち方) & 5. Type (Char: 文字型, Num: 数値型)

**A社**

**B社**

	Identifier (CRF,DB)	Value of Identifier	AESEQ	AESPID
A社	Line No.	1, 2, 3, 4, 5, ...	1, 2, 3, 4, 5, ...	
B社	Line No.	1, 2, 3, 1, 2, 3, ...	1, 2, 3, 1, 2, 3, ...	×
	Page No., Line No.	P101-L01,P101-L02, P101-L03, P102-L01,P102-L02, ...	P101-L01,P101-L02, P101-L03, P102-L01,P102-L02, ...	×
	Page No., Line No.	P101-L01,P101-L02, P101-L03, P102-L01,P102-L02, ...	1, 2, 3, 4, 5, ...	P101-L01,P101-L02, P101-L03, P102-L01,P102-L02, ...

## 6. Controlled Terms / Format

Variable: --STRF, --ENRF

Protocol: ABC123      Page 4  
症例番号: 701      症例報告書

**前治療**

No	薬剤名・療法名
1	アスピリン
2	キニジン
3	

Protocol: ABC123      Page 61  
症例番号: 701      症例報告書

**併用療法**

No	薬剤名・療法名	以前より 継続	開始日	終了日	継続中
1	フェノバルビタール	<input type="checkbox"/>	2005-01-01	2005-03-31	<input type="checkbox"/>
2	ピリドキシン	<input checked="" type="checkbox"/>		2005-12-31	<input type="checkbox"/>
3	トコフェロール	<input type="checkbox"/>	2005-06-15		<input checked="" type="checkbox"/>

**CM** (Concomitant Medications) :

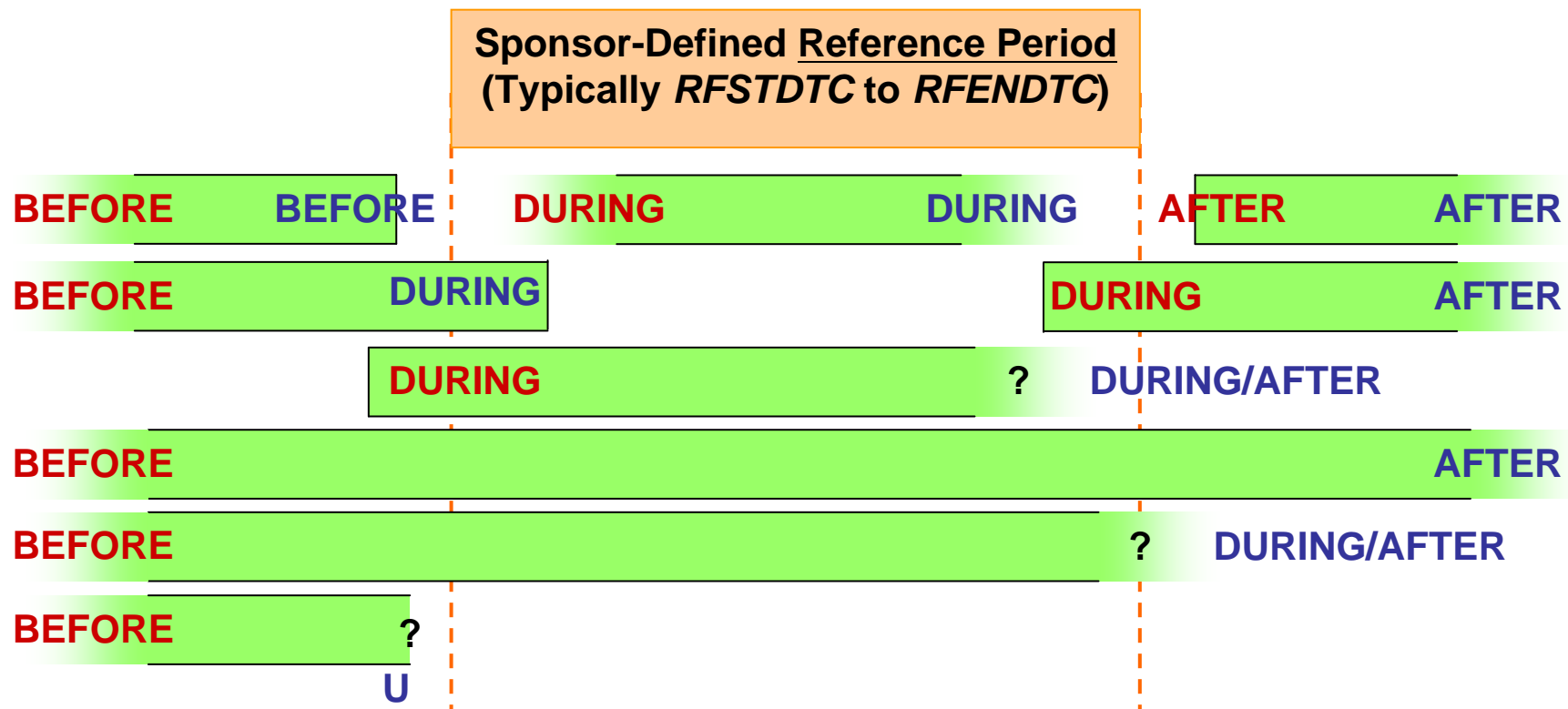
STUDYID	USUBJID	CMSEQ	CMTRT	CMSTDTC	CMENDTC	CMSTRF	CMENRF
ABC123	701	1	Aspirin			BEFORE	BEFORE
ABC123	701	2	Quinidine			BEFORE	BEFORE
ABC123	701	3	Phenobarbital	2005-01-01	2005-03-31		
ABC123	701	4	Pyridoxine		2005-12-31	BEFORE	
ABC123	701	5	Tocopherol	2005-06-15			AFTER

## 6. Controlled Terms / Format

Variable: **--STRF**, **--ENRF**

Value: **“BEFORE”, “DURING”, “AFTER”,  
“DURING/AFTER”, “U”**

~~PREVIOUS  
ONGOING  
CONTINUING  
etc.~~



# 7. Date/Time Values

**ISO8601**: YYYY-MM-DDThh:mm:ss

\* T = separator to indicate time information

	Date and Time as Originally Recorded	Interval of Uncertainty	Nominal Date/Time (--DTC)
1	December 15, 2003 13:14:17	Complete date	2003-12-15T13:14:17
2	December 15, 2003 13:14	Unknown seconds	2003-12-15T13:14
3	December 15, 2003 13	Unknown minutes	2003-12-15T13
4	December 15, 2003	Unknown time	2003-12-15
5	December, 2003	Unknown day	2003-12
6	2003	Unknown month	2003

	Interval of Uncertainty	Nominal Date/Time (--DTC)
1	Between 10:00 and 10:30 on the Morning of December 15, 2003	2003-12-15T10:00/2003-12-15T10:30
2	Between the first and the tenth of December, 2003	2003-12-01/2003-12-10
3	Sometime in the first half of 2003	2003-01-01/2003-06-30

## 8. コメントは1つのDomain「CO」に集約される

### 症例報告書

有害事象		症例番号:23				Page 101
No	事象名	発現日	重症度	転帰	回復日	コメント
1						AAAA
2						BBBB
3						CCCC
ページ コメント	DDDDDDDD					

症例番号:23 Page 201

### 総合コメント

EEEEEEEEEEEEEEEE
EEEEEEEEEEEEEEEE

どのドメインのどのオブザベーションに対するコメントか  
(リレーションをとる変数を用意)

### CO domain

STUDYID	DOMAIN	RDOMAIN	USUBJID	COSEQ	IDVAR	IDVARVAL	COREF	COVAL	COVAL1
AB01	CO	AE	AB01-1-23	1	AESEQ	1	P101	AAAA	
AB01	CO	AE	AB01-1-23	2	AESEQ	2	P101	BBBB	
AB01	CO	AE	AB01-1-23	3	AESEQ	3	P101	CCCC	
AB01	CO	AE	AB01-1-23	4	AEGRPID	101-AE	P101	DDDD	DDD
AB01	CO		AB01-1-23	5			P201	EEEE	EEEE

# SDTM実装例（併用療法: CM [抜粋]）

Protocol: AB 症例番号: 11		症例報告書	併用療法				Page 61
No.	薬剤名・療法名	1回投与量 (mg)	以前より 継続	開始日	終了日	継続中	
1	フェノバルビタール	60	<input type="checkbox"/>	2005/ 01/ 01	2005/ 03/ 31	<input type="checkbox"/>	
2	ピリドキシン	30-60	<input checked="" type="checkbox"/>	/ /	2005/ 12/ 31	<input type="checkbox"/>	
3	トコフェロール	100	<input type="checkbox"/>	2005/ 06/ (?)	/ /	<input checked="" type="checkbox"/>	

CM (Concomitant Medications)								
		Type: Num	Char	Controlled Terms / Format				
USUBJ ID	CM SEQ	CMTRT	CMDOSE	CMDOS TXT	CMSTDTC	CMENDTC	CMSTRF	CMENRF
AB-1-11	1	Phenobarbital	60		2005-01-01	2005-03-31		
AB-1-11	2	Pyridoxine	(45? 60?)	30-60		2005-12-31	BEFORE	
AB-1-11	3	Tocopherol	100		2005-06			AFTER

Core:  
Req:  
Type:  
Num

ISO8601: YYYY-MM-DDThh:mm:ss

BEFORE, DURING,  
AFTER, DURING/AFTER  
U

~~PREVIOUS~~  
~~ONGOING~~  
~~CONTINUING~~

# SDTM vs. ADaM

## SDTM

(Study Data Tabulation Model)

→ 症例データの一覧表

## ADaM

(Analysis Dataset Model)

→ 解析結果

### 年・月・日不明時の日付処理

(欠損を代入した日付変数利用可)

--DTC "2006-07"



--DT "2006-07-(\*\*)" \*\* eg.15  
--DTF "D"

### 時系列データ

(Day 0 設定変数利用可)

--DY Day -3, -2, -1, 1, 2, 3, ...



ANLDY Day -2, -1, 0, 1, 2, 3, ...

### 順序カテゴリカルデータ

(コード化した数値型変数利用可)

Variable Name	Variable Label	Type	Decodes/Format
<b>AESEV</b>	Adverse Event Severity	<b>Char</b>	Mild, Moderate, Severe, Life-threatening



Variable Name	Variable Label	Type	Decodes/Format
<b>AESEVN</b>	Adverse Event Severity <u>Numeric Code</u>	<b>Num</b>	1 = Mild, 2 = Moderate, 3 = Severe, 4 = Life-threatening



# ADaM: Imputation Indicator Variables

## Statistical Analysis Date Variables

	<b><u>ADaM</u></b> Date Variables		<b><u>SDTM</u></b> V3.1 Date-Time Variables
Missing Elements	--DT Date ** <b>Imputed</b>	--DTF	Corresponding --DTC String
<b>None</b>	YYYY-MM-DD	blank	YYYY-MM-DD
<b>Day</b>	YYYY-MM-**	D	YYYY-MM
<b>Month</b> (and Day)	YYYY-**-(**)	M	YYYY
<b>Year</b> (and M, D)	****-(**)-(**)	Y	

ADaM Ver. 2.0  
2/15/2006(Draft)

# ADaM: Statistical Analysis Study Day Variables

**SDTM** If Date  reference date (Typically start of treatment)  
--DY = Date - reference date + 1  
Else if Date  reference date (Typically start of treatment)  
--DY = Date - reference date

**ADaM** ANLDY = Date - reference date + 1

Date	Jan	1	2	3	4	5	6	7	8	9
Visit		SCR				TRT				FU
SDTM	--DY	-4	-3	-2	-1	1	2	3	4	5
ADaM	ANLDY	-3	-2	-1	0	1	2	3	4	5

ADaM Ver. 2.0  
2/15/2006(Draft)

# ADaM: Numeric Code Variables

Variable Name	Variable Label	Type	Decodes/Format	Origin	Role
AESEV	Adverse Event Severity	Char	Mild, Moderate, Severe, Life-threatening	CRF	Qualifier
<b>AESEVN</b>	Adverse Event Severity <u>Numeric Code</u>	<b>Num</b>	<b>1 = Mild, 2 = Moderate, 3 = Severe, 4 = Life-threatening</b>	Coded from AE.AESEV	Selection, Analysis

ADaM Ver. 2.0  
2/15/2006(Draft)

1 . 医薬データの現状と標準化の重要性

2 . CDISCとは？

3 . CDISCのデータ定義を理解する  
(SDTMを中心に)

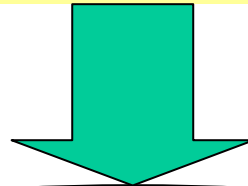
4 . 医薬データの標準化を目指して…

# Need for Standards

臨床試験におけるデータ交換を考えると・・・

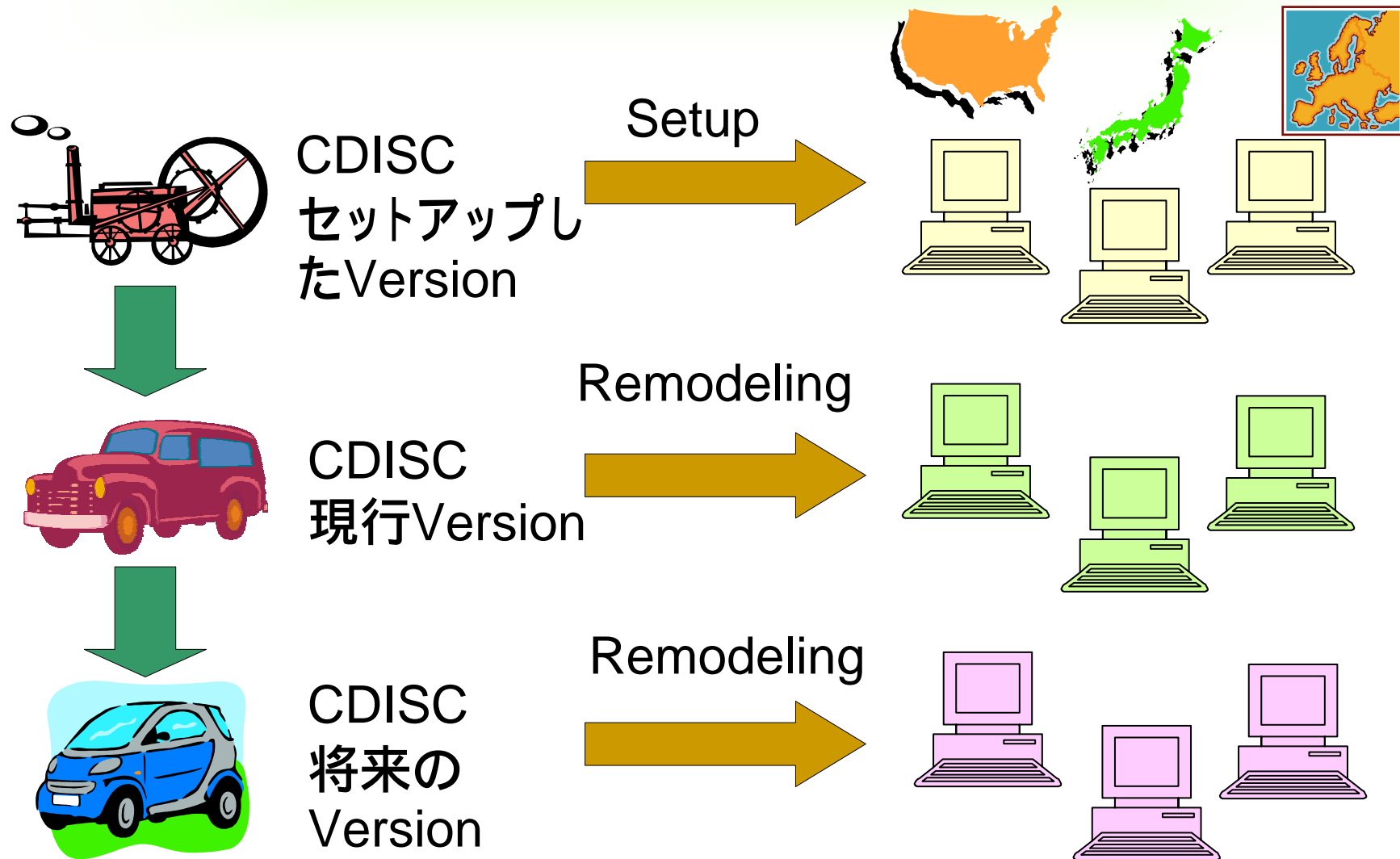
私たちは、データ定義の、真のGlobal Standards  
が欲しい

データの持つ情報(定義・属性等)を損なうことなく  
正しく転送できるStandardsが欲しい

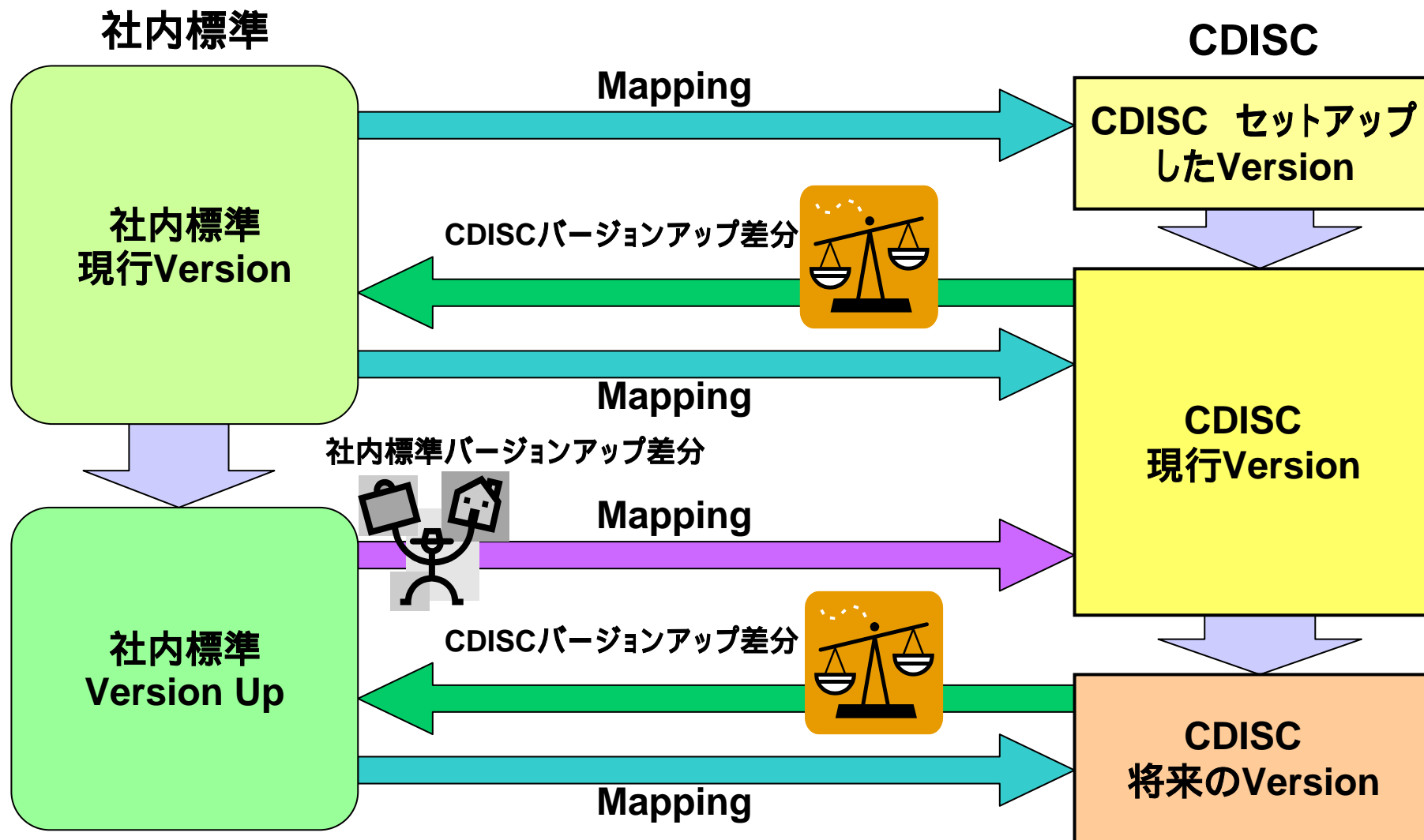


**CDISC**

# CDISCの標準に合わせてセットアップするか・・・



# 社内標準とCDISCの両立: Mappingにて対応か・・・



# Global Standardsが Global Standardsたるには・・・

**将来にわたって大幅な変更のない、確固たるデータ定義**

策定された現行のデータ定義では、個々の機関・企業、自国のデータ定義に合わない

- ➡ 機関・企業、自国のデータ定義をGlobal Standardに合わせられないか
- ➡ Global Standardに機関・企業、自国のデータ定義を反映できないか

**機関・企業、自国におけるデータの定義上の課題や工夫を、  
声を出して標準化策定団体に届ける努力も必要！！**



# まとめ

- データ定義(変数名、変数定義)が確固たるものになっていない状態では、いくら優秀なプログラムでも品質を保持した出力を得られない場合がある。
- 医薬データ加工の効率化と、出力の品質保持のためには、データ定義の標準化が欠かせない。
- CDISCの策定する標準のデータ定義は、医薬データ加工の効率化を考える上で有益なものとなり得る。
- グローバルスタンダードを真に確固たるものにするためには、機関・企業における課題・工夫の声を、標準化策定団体に届ける努力も必要である。

# CDISC: Information & Contacts

## Global

For **standards and information**,  
see <http://www.cdisc.org/>

**eNewsletters** available via e-mail:  
contact Shirley Williams  
swilliams@cdisc.org  
or sign up on the CDISC website.

**Technical questions:**  
Julie Evans jevans@cdisc.org  
or Public Discussion Forum

**Education and Membership:**  
Frank Newby fnewby@cdisc.org

**Rebecca Kush** rkush@cdisc.org

## Japan

Japan CDISC Group (**JCG**)

**Yoshio Tsukada**  
(塚田 良雄)  
cdiscjtsukada@yahoo.co.jp

参加に関してのお問い合わせ

**JCGメンバーシップ担当:**

さかした のぶあき  
**坂下 総亮**  
cdiscjsaka@yahoo.co.jp