IPSEN
Innovation for patient care

CDISC SDTM et Data Management

Isabelle Abousahl
Head of Data Management
Elisabeth Campain-Teulon
Data Warehouse Manager

Groupe des utilisateurs francophones du CDISC

Réunion du 5 Février 2008 – Bagneux / France

Introduction

This presentation reflects my view on the topic, as Head of a Data Management group.

It is not an expert presentation (except the part presented by Elisabeth Campain-Teulon) however the experts will present at the European Interchange Conference in Copenhagen - April 2008 … be present !)

Introduction

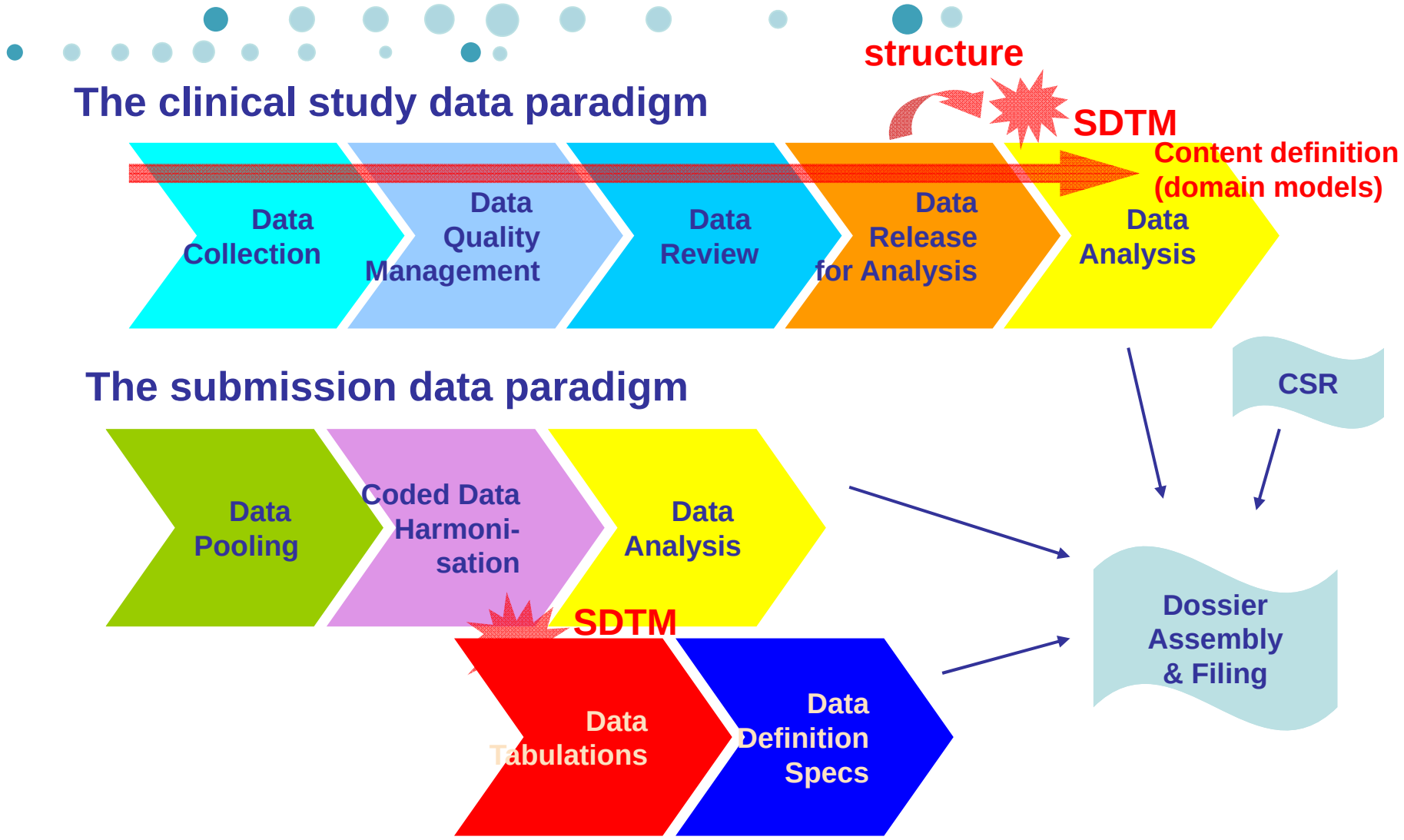
SDTM compliant database – what does it mean for the FDA ?

- Compliance to specifications - checked during the 60 days completeness review by the FDA conducted to accept or refuse the filing
- Traceability of CRF data to SDTM CRT data – checked during site inspection

SDTM compliant database – what does it mean for Data Management ?

.../...

Introduction



IPSEN
Innovation for patient care

Introduction

Data Management role in SDTM implementation ?

Some considerations:

- Data Management may not contribute to SDTM implementation, this would not prevent a company to submit SDTM compliant datasets (ex: DM contracted out to CROs). However, when standardisation is not seen as a global process, it may lead to quality and efficiency issues
- SDTM implementation has to be a cross-functional effort in order to be successful => SDTM implementation by Data Management has to be seen as a contribution, not as a stand-alone mission

Introduction

CDISC standards implementation is fully part of Ipsen Data Management vision and strategy

Ipsen has successfully submitted SDTM compliant databases to the FDA in 2006 and 2007. So our strategy has been successfull up to now, even if there is still a lot to do to improve for future submissions

A vision for Ipsen Data Management

- Aligned on clinical development plan timelines and objectives
- Engaged in quality, not in perfectionism
- Focused on interoperability of: sites, external data sources providers, monitoring teams, partners, PK, drug safety, medical development, statistics and regulatory
- Committed to implementation of low maintenance but smart data systems

Ipsen Data Management strategy

- Adopt, promote and improve process standardisation as a priority in the area where it has the highest impact on quality and/or efficiency
- Realise the investment in EDC and the potential of PRISM clinical data warehouse
- Continue building on Clinical Data Interchange Consortium (CDISC) standards
- Capitalise on the experience and the tools developed for NDA of Lanreotide Autogel in Acromegaly and BLA of Dysport in Cervical Dystonia

IPSEN
Innovation for patient care

Introduction

Data Management interest in SDTM implementation ?

… a last consideration:

- There are commonalities between the data management problems which FDA wants to solve with SDTM and Janus data warehouse and the data management challenges that a Pharma or Biotech company has to face => we can learn from in-depth understanding of the SDTM model for the modelling of our own systems

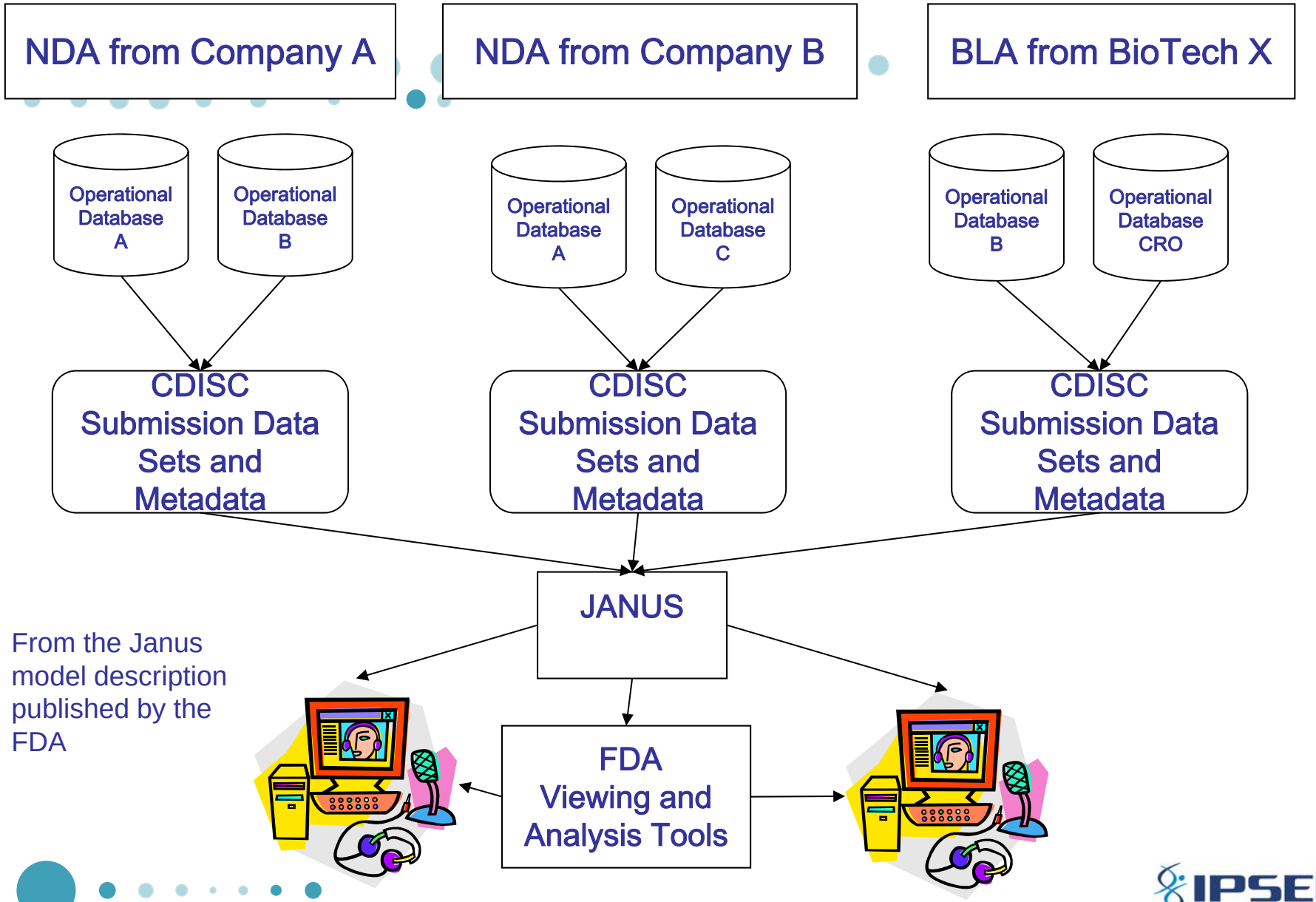
SDTM – Study Data Tabulation Model

*

Data tabulation datasets are one of four ways to represent the human subject Case Report Tabulation (CRT) and equivalent animal data submitted to the FDA. CRTs are also submitted in the format of subject profiles, data listings, and analysis datasets. One benefit to industry of submitting data tabulation datasets that conform to the standard structure is that it minimizes the need to submit the same data in multiple formats.

* From CDISC SDTMIG (version 3.1.2)

FDA Data Warehouse (JANUS)



From the Janus model description published by the FDA

IPSEN
Innovation for patient care

JANUS ←→ SDTM

- JANUS stores many trials
  - Cannot rely on sponsor-supplied keys
- JANUS has fewer tables than datasets
  - One table per observation class
    - Findings
    - Events
    - Interventions
- JANUS has fewer variables
  - No special findings (demographics)
  - No special qualifiers (AE severity)
  - One mechanism to link observations

From the Janus model description published by the FDA

Agenda

What value is Data Management adding in the implementation of CDISC SDTM standards ?

- Leading the process for management of standard CRF libraries in consistency with SDTM content definition
- Modeling the operational database in a pre-SDTM structure ?

Ipsen experience

- Past-submissions and migration tools
- Global Ipsen Data Dictionary

Future plans

Questions & Answers

Data Management & CDISC SDTM

Management of standard CRFs in consistency with SDTM content definition

IPSEN
Innovation for patient care

Guiding principles for standard CRFs

Each Data Domain modelled within standard CRF modules can be:

- Mandatory (ex: Adverse Events)
- Optional (ex: Substance Use)
- Conditional (ex: Breast Cancer History – Oncology standards, is mandatory for Breast Cancer studies)

According to SDTMIG, decision on what data to collect should be based on the scientific objectives of the study rather than SDTM. The CRF module « Mandatory » attribute is based on Ipsen Protocol Template which enforces the collection of some data domains throughout Ipsen studies.

Guiding principles for standard CRFs

The related observation class for each Data Domain modelled within standard CRF modules should be defined (especially for those data sets not already listed in the SDTMIG):

- Events ?
- Interventions ?
- Findings ?

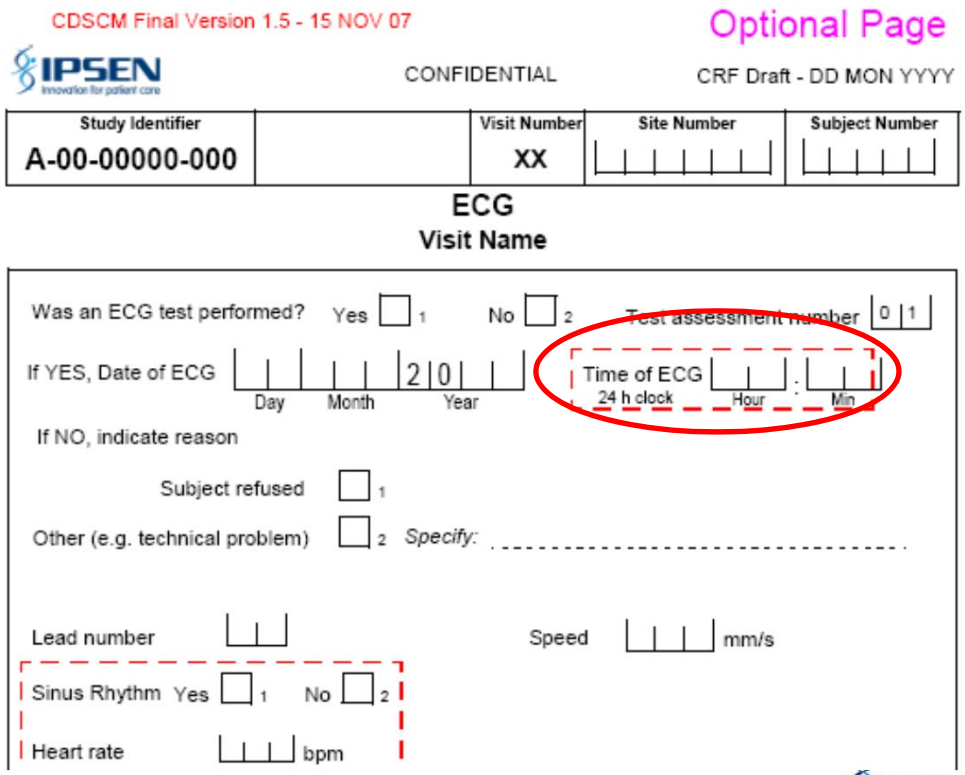
Identifiers should be collected either on each CRF page or at the time the eCRF book is created:

- Study (STUDYID)
- Country (COUNTRY)
- Site (SITEID)
- Investigator (INVID)
- Subject (SUBJID)

Guiding principles for standard CRFs

SDTM core variables categories (required, expected, permissible) should be adressed **directly** or indirectly

This is reflected by the « mandatory/optional » attribute set for each data field in the CRF standard module:



CDSCM Final Version 1.5 - 15 NOV 07

Optional Page

IPSEN Innovation for patient care

CONFIDENTIAL

CRF Draft - DD MON YYYY

Study Identifier		Visit Number	Site Number	Subject Number
A-00-00000-000		XX		

ECG

Visit Name

Was an ECG test performed? Yes ☐ 1 No ☐ 2 Test assessment number 0 1

If YES, Date of ECG Day Month 2 0 Year

Time of ECG 24 h clock Hour : Min

If NO, indicate reason

Subject refused ☐ 1

Other (e.g. technical problem) ☐ 2 *Specify:* .....

Lead number

Speed mm/s

Sinus Rhythm Yes ☐ 1 No ☐ 2

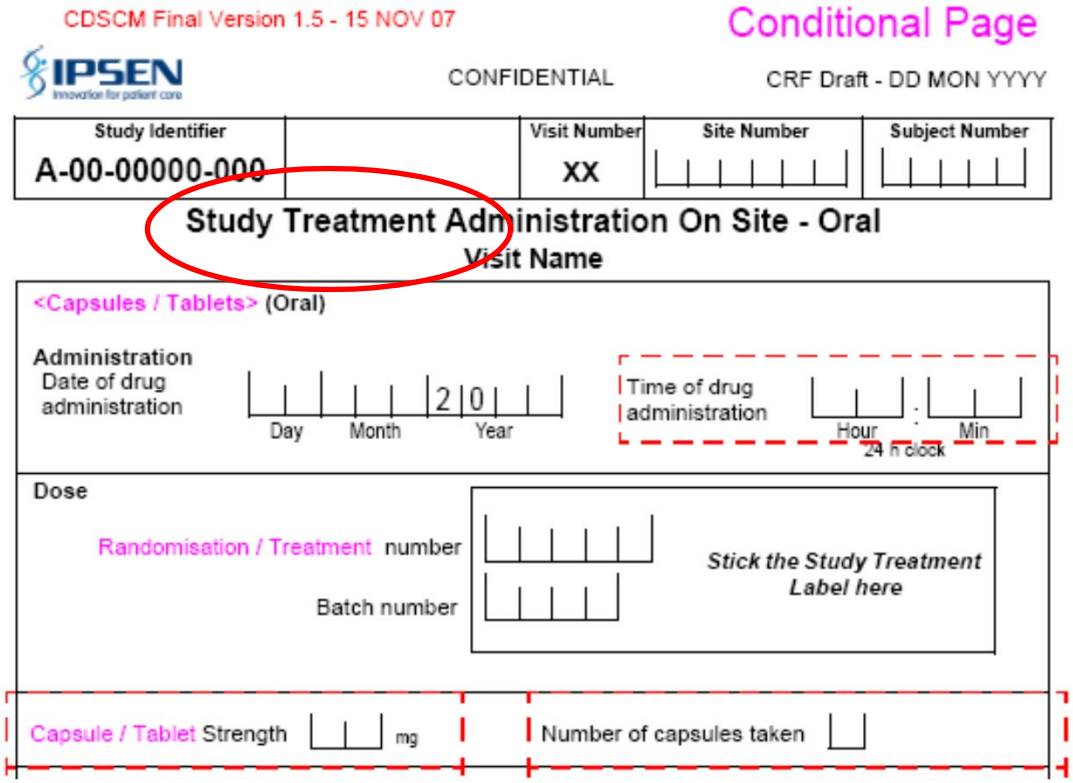
Heart rate bpm

IPSEN Innovation for patient care

Guiding principles for standard CRFs

SDTM core variables categories (required, expected, permissible) should be adressed directly or **indirectly**

SDTM EXPOSURE Domain:
Required variable:
EXTRT
(Name of Actual Treatment)
Do not need to be collected, however, CRF
is designed as per protocol
therefore the study treatment can be indirectly obtained



IPSEN
Innovation for patient care

Guiding principles for standard CRFs

Collection of date and time:

- Day part: DD
- Month part: MMM
- Year part: YYYY
- Time: HH:MM
- Some date and time components may be unknown, or not required (example: day & month part in date of birth may not be collected for data privacy reasons)

Enables standard management of date and time in ISO8601 representation

Management of standard CRFs in consistency with SDTM content definition

Some recommendations:

- Everyone involved in the management of standard and study CRFs should understand the purpose and scope of CDISC standards => training, education
- Study teams should understand the link existing between data acquisition and data submission
- An SDTM expert should be part of the group in charge of elaboration and update of standard CRFs
- Feedback from submission projects may be incorporated in the revision of standard CRFs
- Standard CRFs should be reviewed when a new version of SDTM is released

+ … look at CDASH

Data Management & CDISC SDTM

Modeling the operational database in a pre-SDTM structure ?

IPSEN
Innovation for patient care

SDTM – Study Data Tabulation Model - reminder

* Composed of
  - three general observations classes (interventions, events, findings)
  - other special purposes datasets (demographics, comments, subjects elements, subject visits)
  - the trial design model

Not necessarily to be implemented
in the Data Management operational system

* From CDISC SDTMIG (version 3.1.2)

SDTM – Study Data Tabulation Model - reminder

* Relationships among datasets and records can be represented in different ways:
  - Relating groups of records within a domain (ex: use of - -GRPID to represent a Combination Therapy in CM domain)
  - Relating records in separate datasets (RELREC dataset, ex: AE/CM relationship)
  - Relating non-standard variable values to a parent domain (SUPPQUAL)
  - Relating comments to a parent domain
  - Relating findings observations to events or interventions (use of - - OBJ variable)

The Data Management operational system
may not be able to manipulate easily these relationships

* From CDISC SDTMIG (version 3.1.2)

Guiding principles for a pre-SDTM database ?

Depends on the system used by Data Management:

- Data collection system:
  - CDMS for paper CRF based studies
  - EDC
- Data retrieval system:
  - CDMS
  - SAS environment
  - clinical data warehouse

General considerations

The more differences between version of standards used by the original submitted databases, the more rework and final consolidation steps you will have to perform

- Example: if you need a SAS macro to generate valid ISO8601 date representations in non-standardised studies, you may potentially chose to use this macro also for studies where the ISO8601 format has been derived in the operational database and recompute the valid - - DT field.

General considerations

Front-loading two many computations at the level of the operational database may cause inefficiencies and performance issues

- Example: study day, baseline flag, etc…

General considerations

EDC systems have a data model that is optimised for the performance of data capture and may not be easy to align with the SDTM model

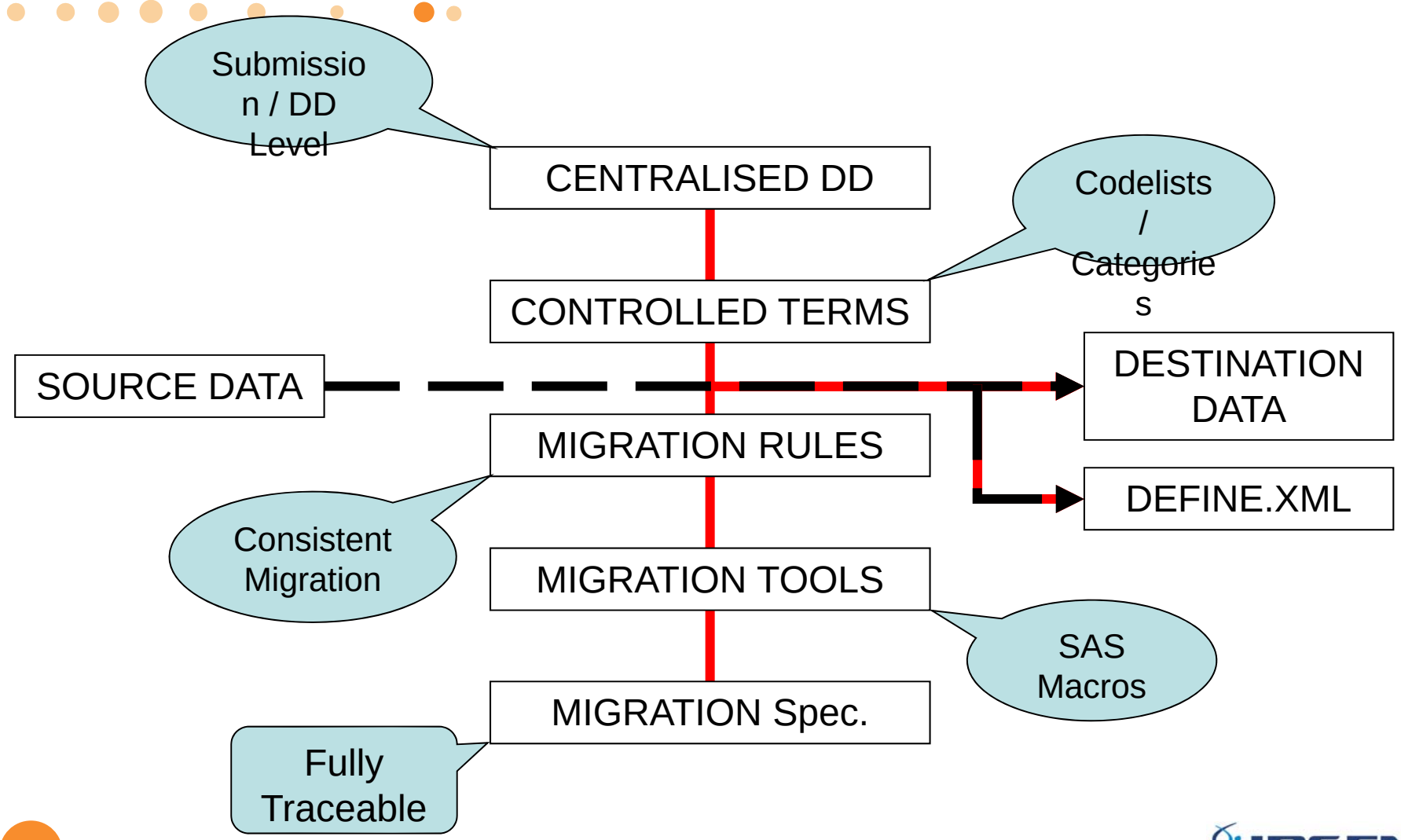
- Example: one itemised underlying database table per data entry form

Ipsen Experience

Past-submissions and migration tools

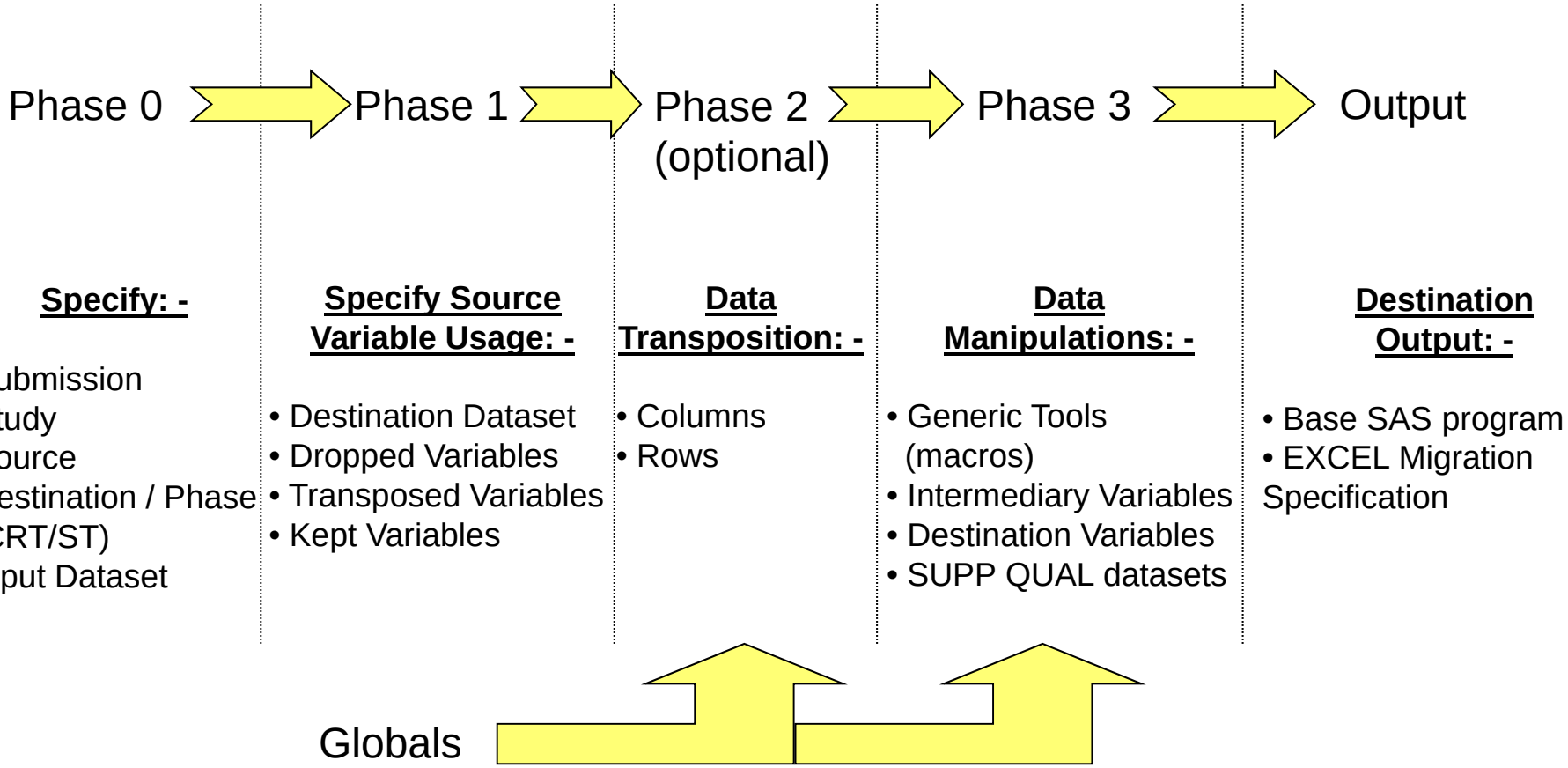
IPSEN
Innovation for patient care

Migration Tool Concepts



IPSEN
Innovation for patient care

Migration Tool



IPSEN
Innovation for patient care

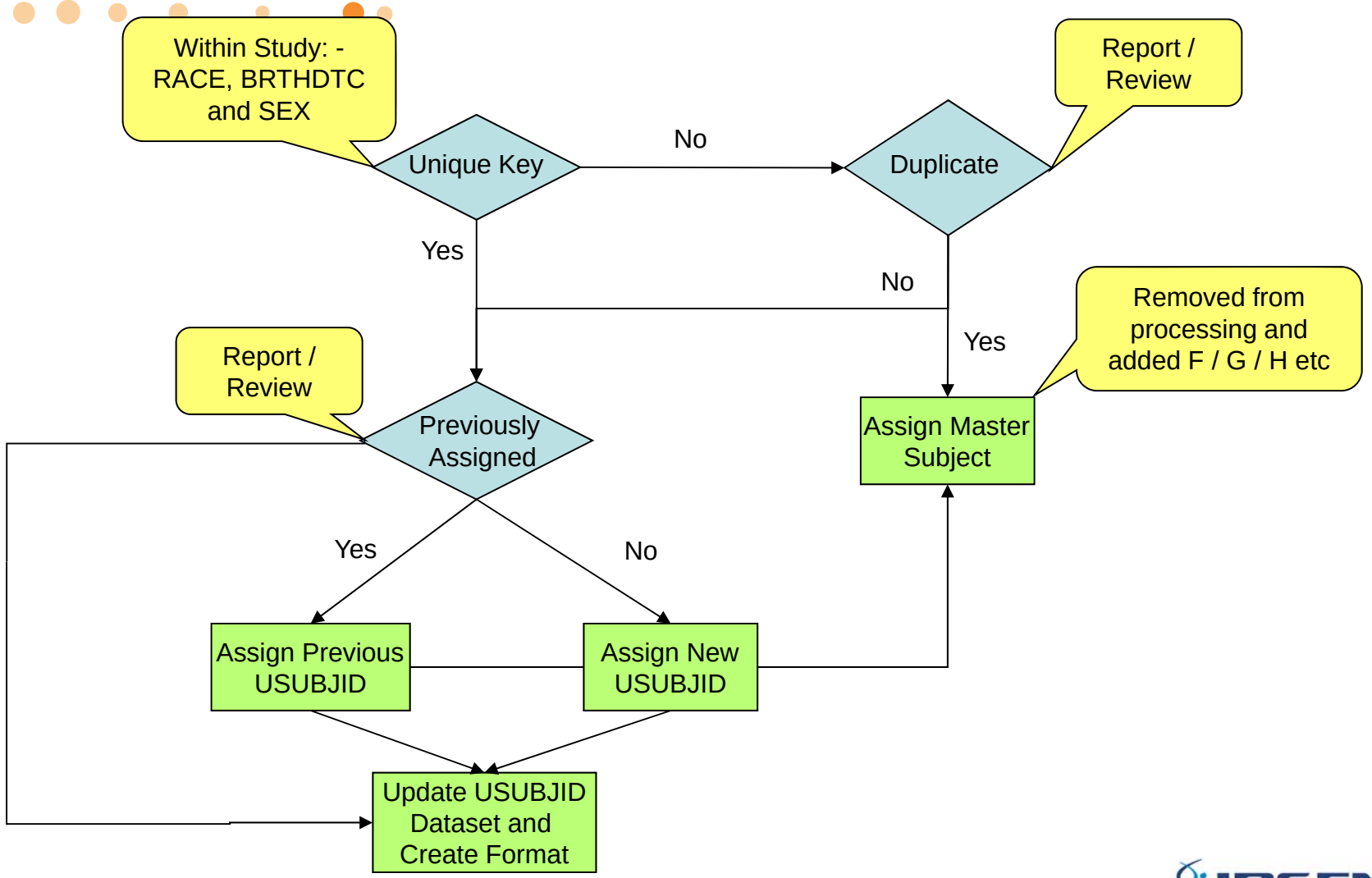
LOOKUP Dataset Structure

Data Migration

DEFINE.XML

MEMNAME	NAME	TYPEC	LENGTH	LABEL	USAGE	CONTRM	SORT_SEQ	ORDER_SEQ	ROLE	ORIGIN	COMMENTS	DDT	DECODE	COMPALG
DM	DOMAIN	C	2	Domain Abbreviation	R	S		2	Identifier	Derived	DM.			
DM	USUBJID	C	8	Unique Subject Identifier	R		3	3	Identifier	Sponsor Defined	See Supplemental Data Definitions Document.			
DM	RFSTDTC	C	20	Subject Reference Start Date/Time	R			5	Timing	Sponsor Defined				
DM	RFENDTC	C	20	Subject Reference End Date/Time	R			6	Timing	Sponsor Defined	Date of last assessment.			
DM	SITEID	C	6	Study Site Identifier	R			7	Record Qualifier					
DM	BRTHDTC	C	20	Date/Time of Birth	IR			10	Result Qualifier	CRF				
DM	AGE	N	8	Age in AGEU at Reference Date/Time	E			11	Result Qualifier					
DM	AGEU	C	6	Age Units	E	S		12	Variable Qualifier	Derived	Defaults to YEARS if AGE is populated.			
DM	SEX	C	1	Sex	R	S		13	Result Qualifier	CRF		Y		
DM	RACE	C	30	Race	E	Y		14	Result Qualifier	CRF		Y		
DM	ETHNIC	C	22	Ethnicity	P			15	Result Qualifier	CRF		Y		
DM	ARMCD	C	8	Planned Arm Code	R	S		16	Result Qualifier	Sponsor Defined			ARM	
DM	ARM	C	40	Description of Planned Arm	R	S		17	Synonym Qualifier	Sponsor Defined	Decode of ARMCD - see label in Value Level Metadata table.			
DM	DMDTC	C	20	Date/Time of Collection	IR			19	Timing	CRF				
DM	DMDY	N	8	Study Day of Collection	IR			20	Timing	Derived				--DY

Supplemental Tools – Example: USUBJID Assignment



IPSEN Innovation for patient care

Supplemental Tools – Other tools

- Compare Team A / Team B
- Quality checks

  Examples:

  All Required variables are present in the Domain
  All Required variables, for all records, are populated in the Domain
  All Expected variables are present in the Domain
  All Expected variables, for at least 1 record, are populated in the Domain
  Unused Variable Check (ensures all legacy variables are reviewed)

- SAS Transfer File Creation (to eCTD area)
- DDT Tool DEFINE.XML Contents

Ipsen Experience

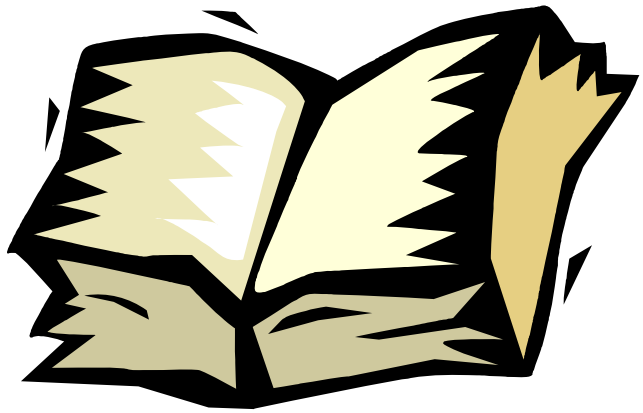
Global Ipsen Data Dictionary (GIDD)

IPSEN
Innovation for patient care

GIDD

What METADATA can we find in GIDD ?

- ✓Tables / Items by therapeutic area
- ✓ Codelists across all therapeutic area
- ✓ Reference lists across all therapeutic area



IPSEN
Innovation for patient care

GIDD Description : tables

GIDD tables summary Report

Current GIDD Version: 4.4

Table Name	Table Label	Table Type	Abbreviation	Area	Mandatory	Optional Reason	Comments
ACSYM	Acromegaly Symptoms	>1 Records/subject/Visit	AC	ENDO	O		
ACSYMCD2	Acromegaly Symptoms Recoding	>1 Records/subject/Visit	AC	ENDO	O		
AEVEN	Adverse Events	>1 Record/Subject	AE	CORE	R		
AEVENCD2	Adverse Events Recoding	>1 Record/Subject	AE	CORE	O	Recoding Only	
AQLQ	Acromegaly Quality of Life Questionnaire	>1 Records/subject/Visit	AQ	ENDO	O		
ASSCA	Acromegaly Symptoms Scale	>1 Records/subject/Visit	AS	ENDO	O		
BIOPS	Biopsy	1 Record/Subject	BR	CORE	O		
CAHIS	Cancer History	>1 Record/Subject	MH	ONCO	R		
CAHISCD2	Cancer History- Recoding	>1 Record/subject	MH	ONCO	R		
CAPPE	Attestation of Completion of Appendices	1 Record/Subject	CA	CORE	M		
CMEDD	Conc. Medication for Studied Disease	> 1 Record/Subject	CM	CORE	O		
CMEDDCD2	Concomitant Medication For Studied Disease Recoding	> 1 Record/Subject	CM	CORE	O	Recoding Only	
COMEN	Comments	>1 Record/Visit/Subject	CO	CORE	O		
COMPL	Drug accountability and Compliance	>1 record/Visit/Subject	CP	CORE	O		
CSENT	Informed Consent	1 Record/Subject	CS	CORE	M		
DEATH	Death Report Form	1 Record/Subject	DE	CORE	O		
DEMOG	Demography	1 Record/Subject	DM	CORE	R		
DEVIA	Deviations	>1 Records/subject/Visit	DV	CORE	M		
EC	Cardiac Echography	>1 record/Visit/Subject	EC	CORE	O		
ECCRF	Echocardiography: CRF part	1 Record/Visit/Subject	EC	CORE	O		
ECG	ECG	>1 record/Visit/Subject	EG	CORE	RC		if results on CRF

GIDD ITEMS : attributes

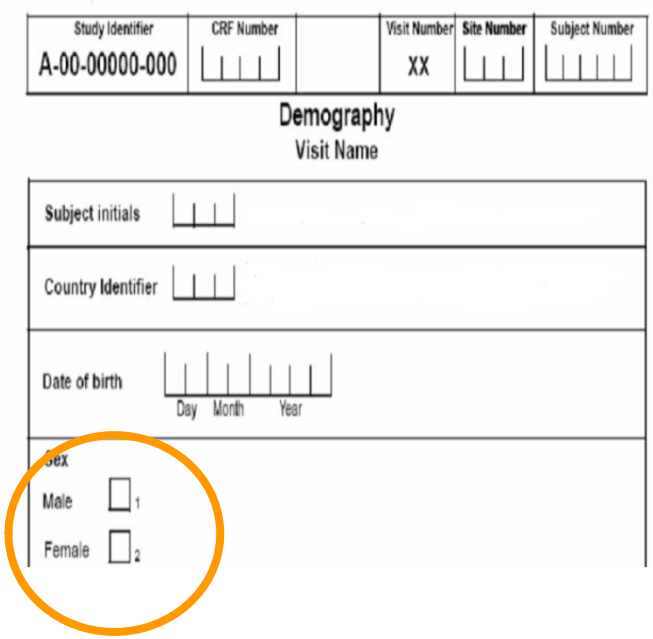
Header

Clinical Tables Details

GIDD Item Name	SDTM Item Name	Domain	Item type	Item Label	SAS Column Type	SAS Format	SAS Column Length	Codelist	Origin	Mandatory
MODCOLCD			Operational	Module Collection Status	N	MODCOL.	3	MODCOL	CRF	M
SUBJINIT		SC	Clinical	Subject Initials	C	$4.	4		CRF	R
COUNTRCD	COUNTRY	DM	Clinical	Country	C	$COUNTR.	3	COUNTR	CRF	R
BRTHDD	BRTHDTC	DM	Clinical	Day of Birth Date	N	2.	3		CRF	R
BRTHMO	BRTHDTC	DM	Clinical	Month of Birth Date	N	2.	3		CRF	R
BRTHYY	BRTHDTC	DM	Clinical	Year of Birth Date	N	4.	4		CRF	R
BRTHDX			Clinical	Date of Birth (Text)	C	$9.	9		DERIVED	R
BRTHDT			Clinical	Date of Birth	N	Date9.	8		DERIVED	R
SEXCD	SEX	DM	Clinical	Sex	N	SEX.	3	SEX	CRF	R
RACECD	RACE	DM	Clinical	Race	N	RACE.	3	RACE	CRF	R
RACEOTH		SC	Clinical	Other Race	C	$200.	200		CRF	M
AGE	AGE	DM	Clinical	Age	N	3	3		DERIVED	R

GIDD: Codelist

CRF :Coded List



GIDD: Codelist

SEX Codelist

Code	Short Label	Long label	SDTM Label
1	**M**	Male	**M**
2	**F**	Female	**F**
3	**B**	Both	**B**

Codelists = list of codes associated with labels

IPSEN Innovation for patient care

GIDD Description: codelist

Codelist terminology in GIDD

Comes from literature when possible

Race : FDA Guideline

Country : ISO 3166

Route : ICH (E2B submission)

Or defined Internally

Lab Test (Hb, HCT, Na,..): Internal Codelist

Or a mix

Action Taken : ICH + Internal codes

GIDD Description: reference list

For each table containing Tests, a reference list has been defined with:

- Standard Units
- Standard Conversion factor (from local to Std unit)
- Standard Significant digits for local and standard units

GIDD Description: reference list

Currently 6 reference lists

Lab
Vital Signs
ECG
Echocardiography
Pharmacokinetics
NCI

GIDD : Reference List

LAB REFERENCE LIST

LBREF TABLE

Test Short Name	TEST NAME	Standard (S.I). UNITS	Standard (SI) UNITS SIGNIFICANT DIGITS	CONVERSION FACTORS -> Div <- Mult	ORIGINAL UNITS	ORIGINAL UNITS SIGNIFICANT DIGITS	DM reference lower limit	DM reference upper limit
LBTESTCD	LBTEST	LBSTRESU	STSDIGIT	CFACTOR	LBORRESU	OUSDIGIT	LBNRLO	LBNRHI
URAC	Uric acid	µmol/L	XX0	59,48	mg/dL	XX.X	120	420
URAC	Uric acid	µmol/L	XX0	5.948	mg/L	XXX	120	420
AXA	Anti Xa activity	IU	X.X				-	-

PCSA Lower Limit	PCSA Upper Limit	Var % decrease	Var % increase	NOMENCL	GROUP	SUB-GROUP	COMMENTS	REFERENCE	Note
LBPCSALO	LBPCSAHI	LBVARDEC	LBVARINC	LBNOMENC	LBGROUP	LBSGROUP	LBCM	LBREF	LBNOTE
-	-	-	50%	13-01	BIOCHEMISTRY	BLOOD		D5, D13, D24, L5, L10, L12, L13, L14, L15	
-	-	-	50%	13-01	BIOCHEMISTRY	BLOOD		D5, D13, D24, L5, L10, L12, L13, L14, L15	
-	-	-	-	05-02	HAEMATOLOGY	HEMOSTASE & COAGULATION		D5	
-	3 x ULN	50%	100%	13-01	BIOCHEMISTRY	BLOOD		D5, D13, D24, L3, L12, L13, L14	

Difference between GIDD and SDTM

Additional tables in GIDD vs SDTM

Example :

GIDD:

- **One lab table per external partner.**

one table for central labs

one table for drug antibodies

- **One table per type of data**

one table for sampling data (CRF/eCRF)

one table for central lab results.

- **SDTM : only one LB Domain**

Difference between GIDD and SDTM

Combined tables in GIDD vs SDTM

Example : Subject initials

- **GIDD : subject initials included with demography data**
- **SDTM : subject initials is described in SC (subject characteristic) DOMAIN**

Difference between GIDD and SDTM

More items in one GIDD table than SDTM domain

- **Operational items**

  Example :Item “status of the page” (used to manage missing pages, not completed pages,..)

- **Yes/No items**

  Example : Any adverse Events ? Any Concomitant medications?

Difference between GIDD and SDTM

More rows with GIDD model than SDTM domain

Example :Inclusion/exclusion criteria

GIDD table : one row per inclusion /inclusion criteria (Yes, No, na)

IE domain : Collect responses to only those criteria that the subject did not meet.

Difference between GIDD and SDTM

GIDD Item name different from SDTM item name.

To facilitate programming, all the coded items contains the SUFFIX= CD

GIDD item name =SEXCD

SDTM item name=SEX

Difference between GIDD and SDTM

GIDD Label item different from SDTM

- *SDTM label are generic*
- *Operational data base : need to be consistent with CRF/eCRF*

Example Medical History

MHSPID item

✓SDTM label= Sponsor ID

✓GIDD label = Row nb

Difference between GIDD and SDTM

GIDD contains code and SDTM contains decode / labels

Example : GIDD sexcd = 1 // MedDRA codes/decode

SDTM sex= M // only MedDRA decode

Difference between GIDD and SDTM

FORMATS

Example : GIDD Sas Formats (Date9. Time5.)

SDTM ISO 8601

Difference between GIDD and SDTM

➢**SUPPQUAL domain**

GIDD : The items are recorded in the main table

SDTM : SUPPQUAL domains are used to record items not defined in the domain model.

➢**RELREC domain**

Example

GIDD : CM table : AE/MH nb item is collected

SDTM : The link must be defined in a RELREC domain

Common points between GIDD and SDTM

- Vertical or normalized structure

Subject	Test	Result
1	HB	Xx
1	HCT	Xx
1	Rbc	xx

- **Horizontal or denormalized structure**

Subject	HB	HCT	RBC
1	XX	XX	XX

Difficulties : Is it the right domain ?

IPSEN case : Local tolerance

Local Tolerance Evaluated Before the 2nd Injection of lanreotide 30 mg PR

The local tolerance should be reported since the last injection of lanreotide 30 mg PR (1st injection).
For each of the following please enter the appropriate information for the patient's symptoms.

	Assessment of the 1st injection of lanreotide 30 mg PR		
	Present?	Length (mm)	Width (mm)
1.Diameter of the induration:	☐ YES ☐ NO	⌊⌋⌋	⌊⌋⌋

The induration size should be determined during palpation by measuring the length and the width from one extremity to the other

	Assessment of the 1st injection of lanreotide 30 mg PR		
Symptom	Symptom Grade*	Time of appearance since the injection	Duration of the symptom between R1 and R2
2.Pain at the injection site:	⌊⌋	⌊⌋⌋⌋ ☐ min. ☐ hrs ☐ days	⌊⌋⌋⌋ ☐ min. ☐ hrs ☐ days
3.Redness:	⌊⌋	⌊⌋⌋⌋ ☐ min. ☐ hrs ☐ days	⌊⌋⌋⌋ ☐ min. ☐ hrs ☐ days
4.Itching:	⌊⌋	⌊⌋⌋⌋ ☐ min. ☐ hrs ☐ days	⌊⌋⌋⌋ ☐ min. ☐ hrs ☐ days

Local tolerance : Event or Finding domain ?

Data Management questions in 2005 :

Issues with Event Domain :

Length and width cannot be mapped in an AE domain (only Finding)

Issues with Finding domain :

The symptoms cannot be coded with MedDra (present in AE domain)

No items to record Duration and delay (present in AE domain)

CDISC answers via forum:

Not clear …can be finding or event ?

Signs and symptoms model should be created…

Local tolerance : Event or Finding domain?

Solution adopted:

Local Tolerance data submitted as a domain from the event observation class: TL and SUPPTL (for length and width data points)

Answers in 2007 : SDTMIG version 3.1.2

2 new domains :

Clinical Events (CE) -EVENTS

Clinical Finding (CF) -FINDINGS

CF domain:	Severity , length and width are collected for each visits using the TEST and TESCD items.
AE domain :	The symptoms are collected with the maximum of severity during the course with duration and delay.

Acromegaly symptoms: Event or Finding domain?

ACROMEGALY SYMPTOMS

For each of the following symptoms please enter the appropriate grade.

SYMPTOMS	GRADE			
	Absent (0)	Mild (1)	Moderate (2)	Severe (3)
1. Headache	☐	☐	☐	☐
2. Excessive perspiration	☐	☐	☐	☐
3. Asthenia	☐	☐	☐	☐
4. Swelling of extremities	☐	☐	☐	☐
5. Joint pain	☐	☐	☐	☐

Acromegaly symptoms: Event or Finding domain?

Ipsen debat in 2005 :

Medical writing : According to the protocol, the Acromegaly symptoms cannot be reported in AE.

Data Management : The AE model is the most appropriate to record these data.

Solution adopted:

Acromegaly symptoms data submitted as a domain in the event observation class (SS)

Acromegaly symptoms: Events or Finding domain?

Answers in 2007 : SDTMIG version 3.1.2

2 new domains :

Clinical Events (CE) -EVENTS

Clinical Finding (CF) -Finding

CE domain: The interest of the domain is to capture clinical events of interest that would not be classified as adverse events.

Future Plans

Innovation for patient care

Future plans

- Impact analysis of SDTM 3.1.2 on Ipsen data standards (being initiated)
- Impact analysis of the standard terminology (to be initiated)
- CDASH to be reviewed and presented at the Clinical Development Data Standard Committee (to be initiated, along with implementation of EDC data standards)
- Integration of migration tools and clinical data warehouse
- Protocol design: prospective versus retrospective creation of the Trial Design Model ?

Questions / Réponses