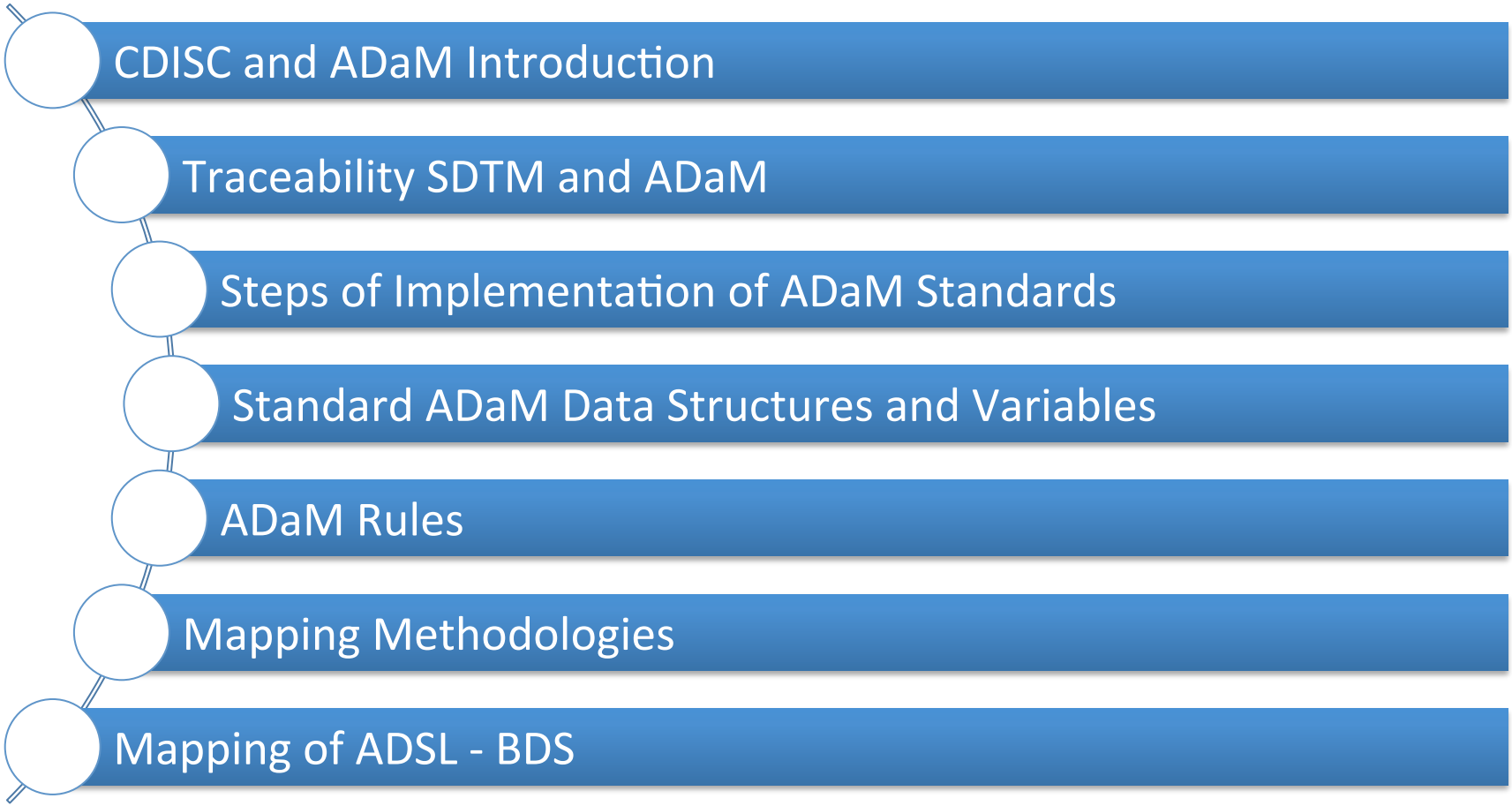
Business & Decision Life Sciences

CDISC Workshop:

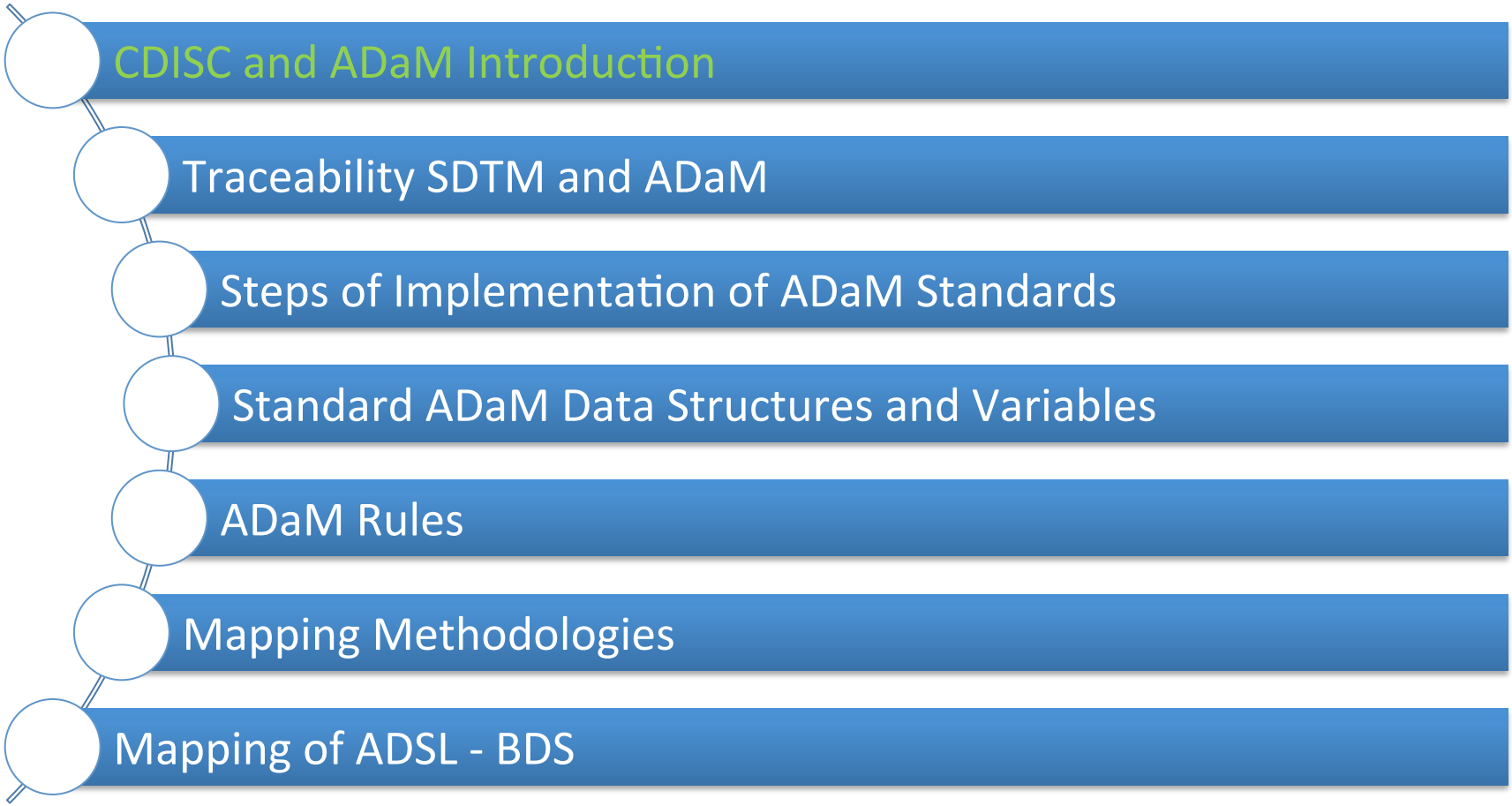
From SDTM to ADaM: Mapping Methodologies

Jessica Minkue Mi Edou / 18th September 2014

From SDTM to ADaM: Mapping Methodologies



From SDTM to ADaM: Mapping Methodologies



CDISC and ADaM Introduction

- CDISC:

  Clinical Data Interchange Standards Consortium

- 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

CDISC and ADaM introduction

CDISC Standards:

- Protocol Representation Model – PRM
- Trial Design Model - TDM
- Clinical Data Acquisition Standards Harmonization – CDASH
- Standards for Exchange of Non-Clinical Data – SEND
- Study Data Tabulation Model - SDTM
- Laboratory Data – LAB
- Analysis Data Model – ADaM
- Terminology
- Glossary
- Define.XML
- Operational Data Model – ODM

CDISC and ADaM introduction

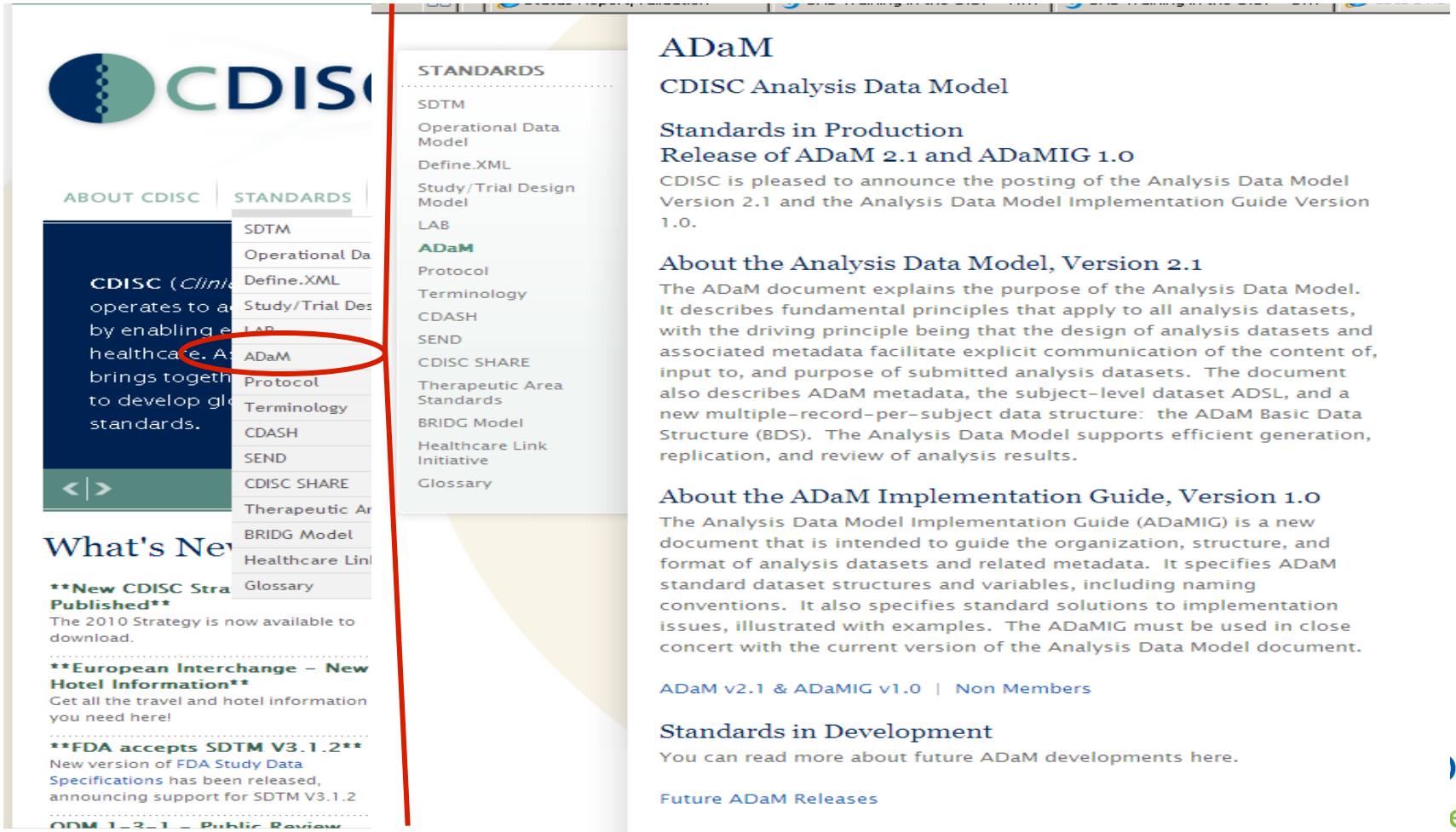
- **CDISC Website**:

- The latest information about CDISC
- Current production standards
- Draft standards (not yet in production)

Business & Decision Life Sciences

CDISC and ADaM introduction

- **CDISC Website - ADaM**:



CDISC and ADaM introduction

- ADaM: Analysis Data Model
- ADaM Team formed in ~ 2001 in recognition that SDTM would not be sufficient for all derived data and analysis
- Fundamental principles and standards to follow in the creation of analysis datasets and associated metadata
- Analysis Datasets (ADaM) are the authoritative source for all data derivations used in statistical analyses
- ADaM represents:
  - Derived data (variables and observations) used for statistical analysis and reporting
  - Observed data (variables and observations) used for statistical analysis and/or traceability
  - Metadata supporting the analysis datasets
- Is developed in addition to SDTM

CDISC and ADaM introduction

SDTM (Study Data Tabulation Model)

- Structure, variables, and variables names are pre-specified
- "Raw Data"
- Data from paper CRFs
- Data from EDC database
- Electronic transfers (labs, etc.)
- Minimal replication of core variables
- Few derived variables
  - Study day, reference dates
- Vertical Finding Structure
- No imputation allowed
- Mostly textual data to facilitate clinical review

ADaM (Analysis Data Model)

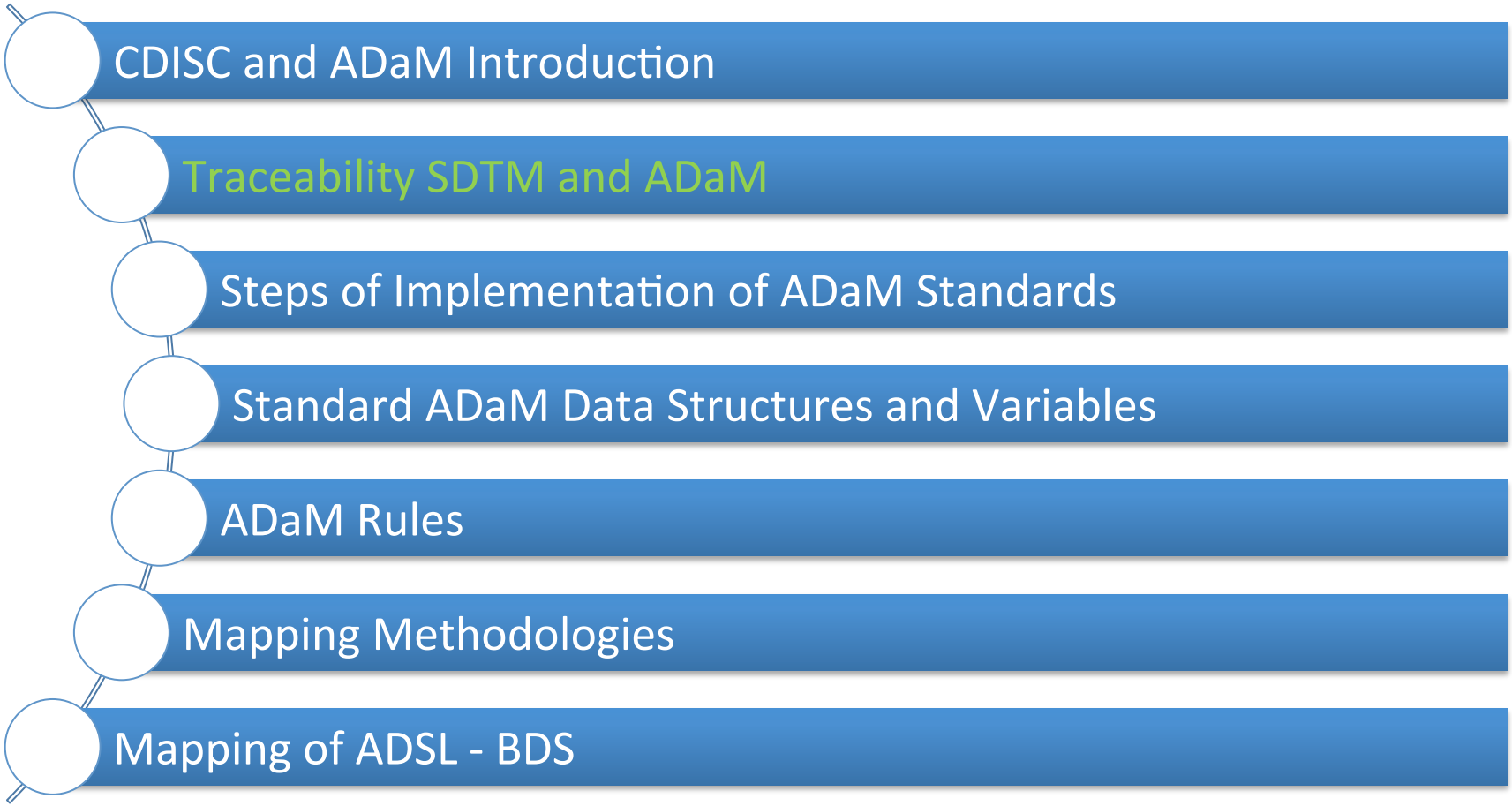
- Can include sponsor-defined derived variables and observations
- Can include SDTM variables and observations
- Can include replication of core variables and other variables that are needed to replicate the analysis with **minimal programming**
- ADaM Subject Level Analysis Dataset (ADSL) is one record per subject
- ADaM Basic Data Structure (BDS) is vertical with flexibility to add columns to support analysis and traceability
- Includes imputations
- Can include numeric values when needed for statistical programming

CDISC and ADaM Introduction

Importance of ADaM

- ADaM provides data that is **ready for analysis** (SDTM is not designed for ease of analysis)
- ADaM helps the reviewer understand:
  - What you said you would do
  - What you did
- TRACEABILITY of data is critical

From SDTM to ADaM: Mapping Methodologies



Traceability SDTM and ADaM

- Understanding relationship between the analysis results, the analysis datasets and the SDTM domains
- Establishing the path between an element and its immediate predecessor
- Two levels:
  - Metadata traceability
    - Relationship of the analysis variable to its source dataset(s) and variable(s)
    - Relationship between an analysis result and analysis dataset(s)
  - Data point traceability
    - Predecessor record(s)

Traceability SDTM and ADaM

Table 1 Demographic Data - Per-Protocol

	Treatment 1	Treatment 2
Baseline body mass index (BMI) [kg/m**2]		
N	167	167
Mean	29.08	29.04
SD	4.84	4.80
Min	20.3	16.0
Median	28.69	28.47
Max	40.1	41.2
Baseline BMI (categorical) [N (%)]		
<25 kg/m**2	41 (24.6%)	71 (21.1%)
25-<30 kg/m**2	60 (35.9%)	130 (38.7%)
>=30 kg/m**2	66 (39.5%)	135 (40.2%)

- **Patient Demographics – Part I**

Patient X5 Page 4 (Demo_V1a for Visit 1a) Page 1 of 1.

Visit Date 11-Dec-2007 Blank Comment

PATIENT DEMOGRAPHICS - Part I **DSCAT = "PROTOCOL MILESTONE"**

Informed consent was obtained on **DSTERM / DSDECOD** **DSSTDTC**

Gender **SEX** 1 = male, 2 = female

Date of birth **BRTHDTC** Age **AGE** years **AGEU**

(Age is automatically calculated when screen is saved and closed)

Height cm

Weight kg

Waist circumference cm

VSORRES / VSORRESU where VSTESTCD = "HEIGHT", "WEIGHT", "WAIST"

	STUDYID	USUBJID	SUBJID	BMI	BMIGR1	BMIGR1N	BMIGR2	BMIGR2N
2	9999-0001	9999-0001-000001	000001	27.777777778	<30 kg/m**2	1	25-<30 kg/m**2	2
3	9999-0001	9999-0001-000002	000002	25.503615702	<30 kg/m**2	1	25-<30 kg/m**2	2
4	9999-0001	9999-0001-000003	000003	26.175194521	<30 kg/m**2	1	25-<30 kg/m**2	2
5	9999-0001	9999-0001-000004	000004	35.15625	>=30 kg/m**2	2	>=30 kg/m**2	3
6	9999-0001	9999-0001-000005	000005	30.968858131	>=30 kg/m**2	2	>=30 kg/m**2	3
7	9999-0001	9999-0001-000006	000006	39.697163916	>=30 kg/m**2	2	>=30 kg/m**2	3
8	9999-0001	9999-0001-000007	000007	25.826446281	<30 kg/m**2	1	25-<30 kg/m**2	2
9	9999-0001	9999-0001-000008	000008	30.103806228	>=30 kg/m**2	2	>=30 kg/m**2	3
10	9999-0001	9999-0001-000009	000009	32.280962683	>=30 kg/m**2	2	>=30 kg/m**2	3
11	9999-0001	9999-0001-000010	000010	28.876133787	<30 kg/m**2	1	25-<30 kg/m**2	2
12	9999-0001	9999-0001-000011	000011	29.372397383	<30 kg/m**2	1	25-<30 kg/m**2	2
13	9999-0001	9999-0001-000012	000012	26.714852608	<30 kg/m**2	1	25-<30 kg/m**2	2
14	9999-0001	9999-0001-000013	000013	32.718619869	>=30 kg/m**2	2	>=30 kg/m**2	3
15	9999-0001	9999-0001-000014	000014	28.715723183	<30 kg/m**2	1	25-<30 kg/m**2	2
16	9999-0001	9999-0001-000015	000015	32.270420377	>=30 kg/m**2	2	>=30 kg/m**2	3

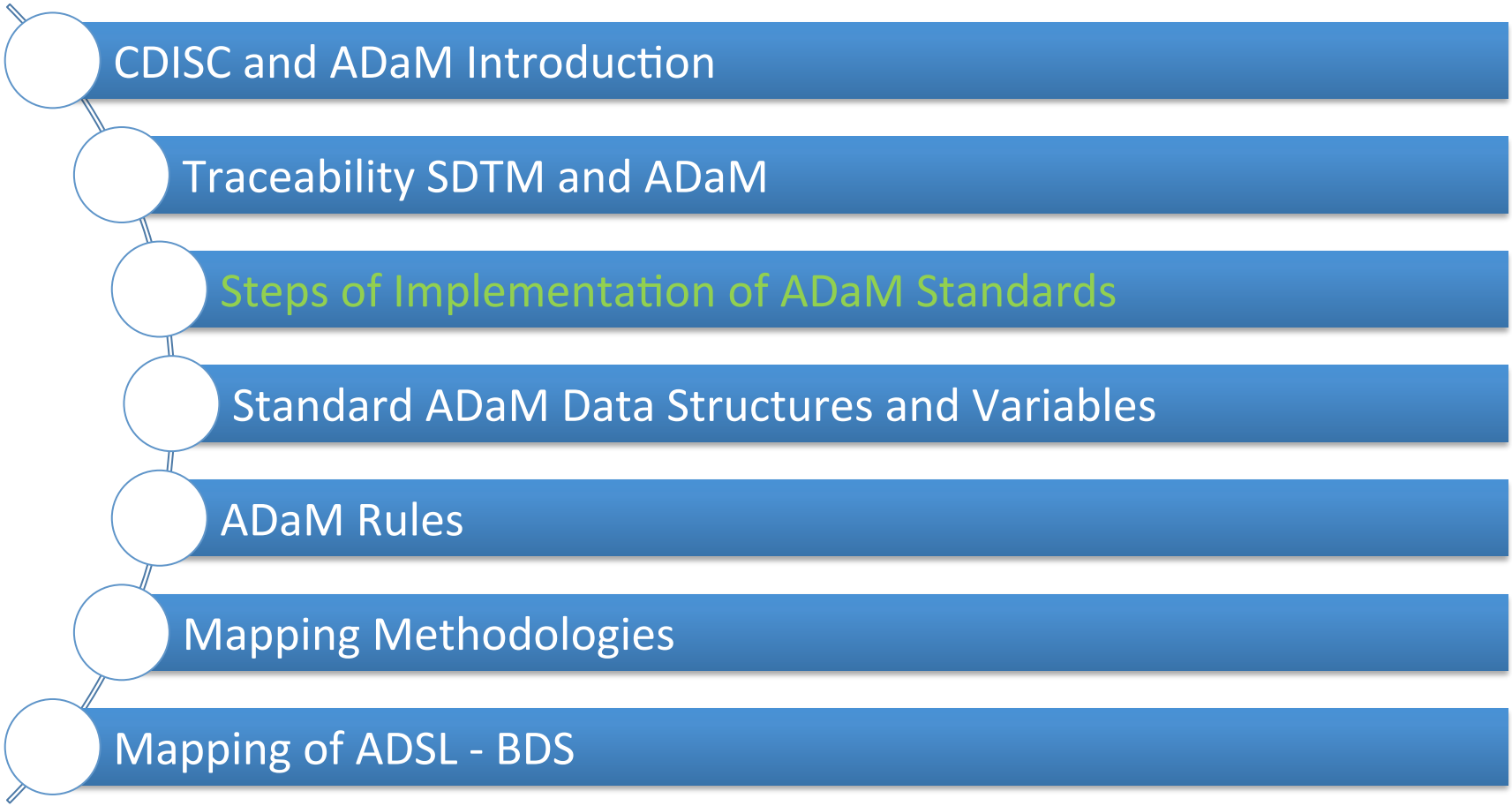
Vital Signs Dataset (VS) ../Tabulations/vs.xpt

Variable	Label	Type	Controlled Terminology	Origin	Role	Comment
STUDYID	Study Identifier	text		Protocol, CRF Page 1	IDENTIFIER	The STUDYID variable has a fixed format: 'XXXX-YYYY', where 'XXXX' indicates the 4-digit compound code and the 'YYYY' the 4-digit study code
DOMAIN	Domain Abbreviation	text	DOMAIN	Assigned	IDENTIFIER	
USUBJID	Unique Subject Identifier	text		Protocol	IDENTIFIER	The USUBJID variable has a fixed format: 'XXXX-YYYY-ZZZZZZ', where 'XXXX' indicates the 4-digit compound code, 'YYYY' the 4-digit study code and 'ZZZZZZ' the 6-digit patient code
VSSEQ	Sequence Number	integer		Derived	IDENTIFIER	Sequence number (automatically generated) to ensure uniqueness within a dataset for a subject
VSTESTCD	Vital Signs Test Short Name	text		Assigned	TOPIC	
VSTEST	Vital Signs Test Name	text		Derived	SYNONYM QUALIFIER	
VSPOS	Vital Signs Position of Subject	text	POSITION	CRF Page 13	RECORD QUALIFIER	
VSORRES	Result or Finding in Original Units	text		Derived, CRF Page 9, 13	RESULT QUALIFIER	
VSORRESU	Original Units	text	VSRESU	CRF Page 9, 13	VARIABLE QUALIFIER	
VSSTRESC	Character Result/Finding in Std Format	text		Derived	RESULT QUALIFIER	
VSSTRESN	Numeric Result/Finding in Standard Units	float	3.1	Derived	RESULT QUALIFIER	

Subject-Level Analysis Dataset Dataset (ADSL) ../Analysis/adsl.xpt

Variable	Label	Type	Controlled Terms or Format	Computational Algorithm or Method	Origin	Role	Comment
STUDYID	Study Identifier	text			DM	IDENTIFIER	The STUDYID variable has a fixed format: 'XXXX-YYYY', where 'XXXX' indicates the 4-digit compound code and the 'YYYY' the 4-digit study code
USUBJID	Unique Subject Identifier	text			DM	IDENTIFIER	The USUBJID variable has a fixed format: 'XXXX-YYYY-ZZZZZZ', where 'XXXX' indicates the 4-digit compound code, 'YYYY' the 4-digit study code and 'ZZZZZZ' the 5-digit patient code
SUBJID	Subject Identifier for the Study	text			DM	IDENTIFIER	
HEIGHT	Baseline Height (cm)	integer		ADSL.HEIGHT	Derived	ANALYSIS	
WEIGHT	Baseline Weight (kg)	integer		ADSL.WEIGHT	Derived	ANALYSIS	
BMI	Baseline BMI (kg/m**2)	integer		ADSL.BMI	Derived	ANALYSIS	
BMIGR1	Category 1 of Baseline BMI	text	BMIGR1L	ADSL.BMIGR1	Derived	ANALYSIS	
BMIGR1N	Category 1 of Baseline BMI, (N)	integer	BMIGR1N	ADSL.BMIGR1N	Derived	ANALYSIS	
BMIGR2	Category 2 of Baseline BMI	text	BMIGR2L	ADSL.BMIGR2	Derived	ANALYSIS	
BMIGR2N	Category 2 of Baseline BMI, (N)	integer	BMIGR2N	ADSL.BMIGR2N	Derived	ANALYSIS	
BMIGR3	Category 3 of Baseline BMI	text	BMIGR3L	ADSL.BMIGR3	Derived	ANALYSIS	
BMIGR3N	Category 3 of Baseline BMI, (N)	integer	BMIGR3N	ADSL.BMIGR3N	Derived	ANALYSIS	

From SDTM to ADaM: Mapping Methodologies



Steps of Implementation of ADaM Standard

Creation of Specifications (mapping)

- Specifications of the analysis datasets structure and computational algorithms. The mapping is used to:
  - Write programs that create ADaM datasets (Data)
  - Generate Define.xml (Metadata)

Table Metadata

Dataset Name
Description
Structure
Purpose
Keys
Location

Variable Metadata

Variable Name
Variable Label
Type
Controlled Terminology or Format
Computational Algorithm or Method
Origin
Role
Comment

Value Level Metadata

Source Variable
Value
Label
Type
Controlled Terminology
Origin
Role
Comment

Controlled Terminology

Code Value
Code Text

Computational Algorithms

Reference Name
Computational Method

Steps of Implementation of ADaM Standard

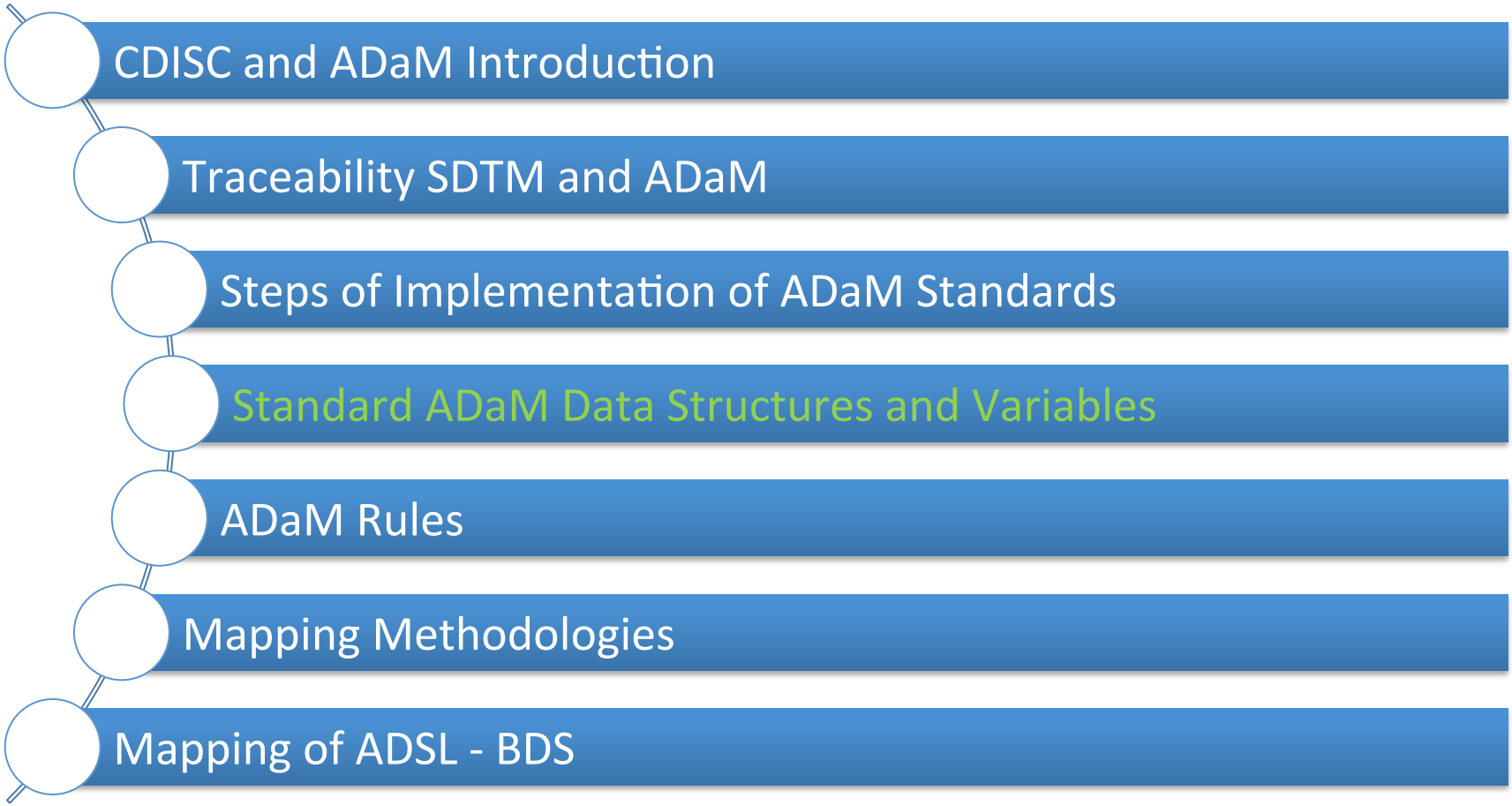
Creation of ADaM Datasets and Define.xml

- ADaM datasets and Define.xml have to be created following the specifications

Check of ADaM Datasets and Define.xml

- Free tools: Open CDISC
- Commercial tools also exist

From SDTM to ADaM: Mapping Methodologies



Standard ADaM data structures and variables

Subject Level Structure (ADSL):

- Reserved dataset name 'ADSL'
- One record per subject, regardless of study design
- Used as a source for variables required in other datasets and denominator values for population of interest
- Used to support tables such as subjects dispositions, subject demographics, baseline characteristics
- Not intended to be the only file that supports all subject level analyses
- It is required in a CDISC-based submission of data from a clinical trial

Standard ADaM data structures and variables

The ADaM Basic Data Structure (BDS):

- Designed with the majority of analysis files in mind.
- Can be loosely described as one or more records
  per subject, (USUBJID)
  per analysis parameter, (PARAM)
  per analysis time point (as appropriate) (AVISIT)
- Includes observations for both observed and derived values required for analysis
- "Analysis-focused" design: variables on the record are focused on the analysis variables, i.e. PARAM/AVAL.
- Allows flexibility to add supportive columns as needed

Standard ADaM data structures and variables

ADAE Structure:

- General data structure and content typically found in analysis datasets used for common safety analysis of adverse events

ADaM Time to Event (TTE) Structure:

- BDS structure plus additional TTE variables
- Should be named following the ADaM standard naming convention (not necessary ADTTE)
- It is sponsor's decision to determine how many ADTTE datasets would be adequate for a given study

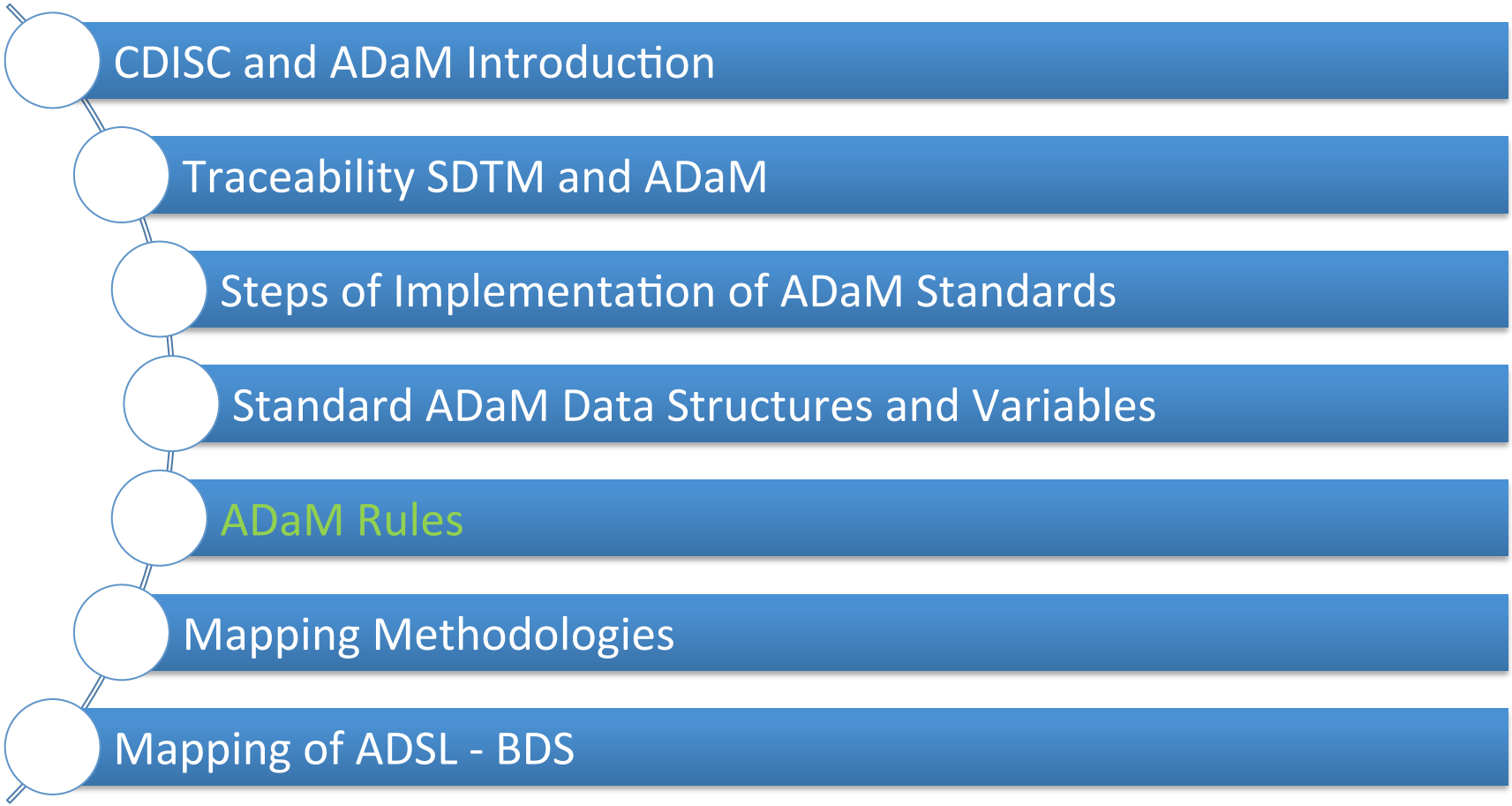
Standard ADaM data structures and variables

Variables core attributes:

- Required: must be included in the dataset
- Conditional: The variable must be included in the dataset in certain circumstances
- Permissible: The variable may be included in the dataset, but is not required.

Unless otherwise specified, all ADaM variables are populated as appropriate, meaning nulls are allowed.

From SDTM to ADaM: Mapping Methodologies



ADaM Rules

General Datasets Naming Conventions

- At a minimum, one analysis dataset is required - ADSL
- Analysis datasets must follow the naming convention ADxxxxxx

ADaM Rules

General Variables Naming Conventions

- ADaM variables names must be no more than 8 characters in length, start with a letter (not underscore) and be comprised only of letters (A-Z), underscore (_) and numerals (0-9).
- ADaM variables labels must be no more than 40 characters in length. All ADaM character variables must be no more than 200 characters in length.
- If an SDTM variable appears in an ADaM dataset then all attributes must remain the same (same name, same meaning, same values)
- ADaM standard names must be used when a variable contains the content defined in the model
- In a pair of corresponding variables (e.g TRTP and TRTPN, AVAL and AVALC), the primary or most commonly used variable does not have the suffix or extension (e.g N for Numeric or C for Character).

ADaM Rules

General Variables Naming Conventions

- Lower case letters "xx", "y" and "zz" that appear in a variable name or label must be replaced as follows:
  - The letters "xx" in a variable name (e.g , TRTxxP) refers to a specific period where "xx" is replaced with a zero-padded two-digit integer [01-99].
  - The lower case "y" in a variable name (e.g SITEGRy) refers to a grouping or other categorization, an analysis criterion, or an analysis range, and is replaced with a single digit [1-9].
  - The lower case "zz" in a variable name (e.g, ANLzzFL) is an index for the zzth record selection algorithm where "zz" is replaced with a zero-padded two digit integer [01-99].
- When an asterisk (*) is used as a variable name prefix or suffix (e.g, *STM) it must be replaced by a suitable character string
- The names of date imputation flag variables end in DTF, and the name of time imputation variables end in TMF.
- The name of all other character flag (or indicator) variables end in FL, and the names of the corresponding numeric flag (or indicator) variables end in FN. If the flag is used, the character version is required but the numeric version can also be included.
  - e.g: The character flag ITTFL and it numeric version ITTFN

ADaM Rules

General Variables Naming Conventions

- IF any combining of the SDTM character categories is done, the name of the derived ADaM character grouping variable should end in Gry and the name of the numeric equivalent should end in GRyN where y is an integer from 1-9 representing a grouping scheme. Truncation of the original variable name may be necessary when appending suffix fragments Gry, or GRyN.

- Variables whose names end in GRy are grouping variables, where "y" refers to the grouping scheme or algorithm
  - e.g: SITEGR3 is the name of a variable containing site group names, where the grouping has been done according to the third site grouping algorithm; SITEGR3 does not mean the third group of sites.

- In general, if SDTM character variables are converted to numeric variables in ADaM datasets, then they should be named as they are in SDTM with an "N" suffix added. If necessary, to keep within the 8-character variable name length limit, the last character may be removed prior to appending the N.
  - Note: it applies only to numeric variables whose values map one-to-one to the values of the equivalent character variables. This convention does not apply to date/time variables.

ADaM Rules

General Timing Variables Conventions

- Variables whose names end in DT are numeric dates.
- Variables whose names end in DTM are numeric datetimes.
- Variables whose names end in TM are numeric times.
- If a *DTM and associated *TM variable exist, then the *TM variable must match the time part of the *DTM variable. If a *DTM and associated *DT variable exist, then the *DT variable must match the date part of the *DTM variable.
- Variables whose names end in DY are relative day variables. In ADaM and SDTM there is no day 0.

ADaM Rules

General Timing Variables Conventions

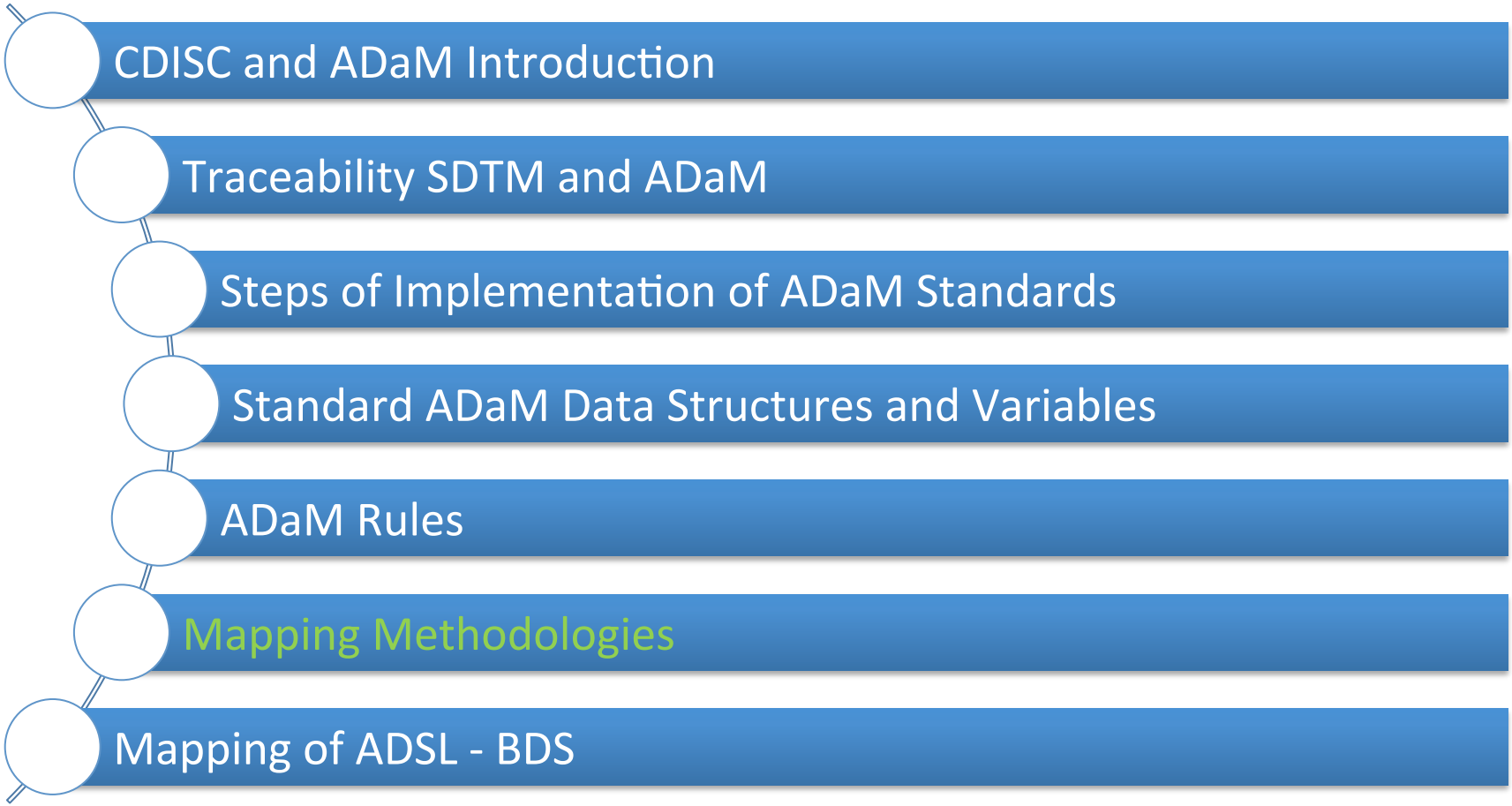
- Variables whose names end in DTF are date imputation flags. Can have the values Y if the whole date is imputed, M if month and day are imputed or D if only the day is imputed.
- Variables whose names end in TMF are time imputation flags. Can have the values H if the entire time is imputed, M if minutes and seconds are imputed or S if only seconds are imputed.
- Names of timing start variables end with an S followed by the two characters indicating the type of timing (e.g: SDT, STM).
- Names of timing end variables end with an E followed by the two characters indicating the type of timing (e.g: EDT, ETM).

ADaM Rules

General Flag Variable Conventions

- Population flags must be included if the dataset is analyzed by the given population. At least one population flag is required for datasets used for analysis. All applicable subject level population flags must be present in ADSL.
- Character and numeric subject-level population flag names end in FL and FN, respectively. Parameter-level population flags end in PFL and PFN, and record-level population flag end in RFL and RFN.
- For subject-level character population flag variables; N=no (not included in the population), Y=yes (included). Null values are not allowed.
- For subject-level numeric population flag variables; 0=no (not included in the population), 1=yes (included). Null values are not allowed.
- For parameter-level and record-level character population flag variables; Y=yes (included). Null values are allowed.
- For parameter-level and record-level numeric population flag variables; 1=yes (included). Null values are allowed.

From SDTM to ADaM: Mapping Methodologies



Mapping Methodologies

Material for an ADaM mapping

- SAP
- PROTOCOL
- SDTM Mapping
- SDTM datasets
- Case Report Form (CRF)
- Sponsor's specifications
- CDISC ADaM implementation guide

Mapping Methodologies

Table Metadata

Dataset Name
Description
Structure
Purpose
Keys
Location

Variable Metadata

Variable Name
Variable Label
Type
Controlled Terminology or Format
Computational Algorithm or Method
Origin
Role
Comment

Value Level Metadata

Source Variable
Value
Label
Type
Controlled Terminology
Origin
Role
Comment

Controlled Terminology

Code Value
Code Text

Computational Algorithms

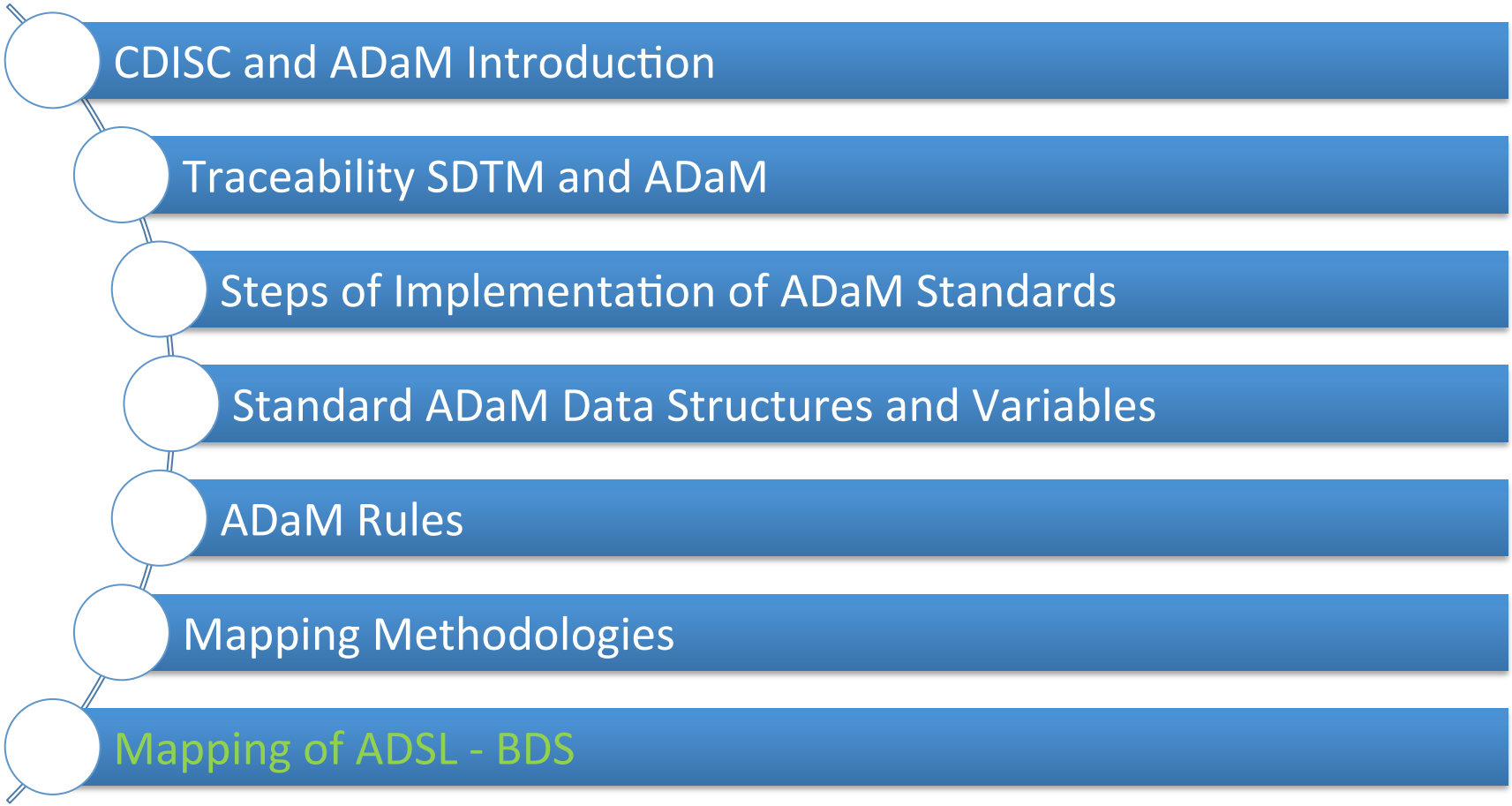
Reference Name
Computational Method

Business & Decision Life Sciences

Mapping Methodologies

- Identify in the SAP and Protocol all the variables and datasets needed for the analysis
- Identify in the ADaM implementation guide all the variables needed for the specific dataset. Pay attention to the conditionnally required variables.
- Use the annoted CRF to trace back the collected information to the SDTM mapping
- Create the computational algorithms, codelist,
- Mapping checks
  - All Required and Computational variables are mapped
  - The proper ADaM standards names are used
  - The proper labels and attributes are used
  - Copied variables from SDTM have the same name, values and attributes
  - The codelist have the correct name and values

From SDTM to ADaM: Mapping Methodologies



Mapping of ADSL

What goes into ADSL

- Required variables
  - Study and subject identifiers (STUDYID,USUBJID)
  - Subject demographics (AGE, RACE, …)
  - Population indicators (IITFL, …)
  - Treatment variables (TRT01P, …)
  - Trial dates (conditionally required)

- Additional variables
  - Baseline characteristics
  - Numeric equivalents of flags
  - Stratification variables
  - Treatment duration and compliance variables
  - Other key visit dates and durations
  - Protocol specific event information, such as death/survival

Example ADSL

Mapping of a BDS

BDS Variables – Study and Subject Identifiers

- Variables that provide information about individual subjects, e.g.,
  - STUDYID: study identifier (req)
  - USUBJID: unique subject identifier (req)
  - SITEID: unique site identifier (perm; required in ADSL)
  - SUBJID: subject identifier used within study (perm; required in ADSL)
  - SITEGRy: pooled site group y (perm)
- Include those that are applicable to the analysis
- All BDS datasets must contain the SDTM STUDYID and USUBJID variables

Mapping of a BDS

BDS Variables – Analysis Parameter

- Variables to describe what is being analyzed
- Required:
  - PARAM: parameter description
    - Uniquely describes analysis variable AVAL
    - Its value should be what appears in statistical tables
    - Include units, specimen type, position, etc. as appropriate
    - Not restricted to 40 characters (but < = 200)
  - PARAMCD: 8 character version of PARAM
- Permitted:
  - PARAMN: numeric version of PARAM
  - PARAMTYP: indicator that parameter is derived
  - PARCATy: categorization of PARAM

Mapping of a BDS

BDS Variables - Analysis Values

- Variables that contain character or numeric values which are used for analysis
  - AVAL: numeric analysis value described by PARAM
  - AVALC: character analysis value described by PARAM
  - AVALCATy: categorization of AVAL and/or AVALC
    - Does not have to be a 1:1 correspondence with AVAL and/or AVALC.

- At least one of AVAL or AVALC is required in BDS

Mapping of a BDS

BDS Variables – More Analysis Variables

- BASE: baseline value (cond)
- BASEC: character baseline value (perm)
- CHG: change from baseline (perm)
- CHGCATy: categorical representation of change from baseline (perm)
- PCHG: percent change from baseline (perm)
- SHIFTy: function of defined pairs such as BASECATy and AVALCATy or BASE and AVAL
  - e.g., shift from Baseline - 'Normal to Abnormal'

Mapping of a BDS

BDS Variables – Timing Variables

There are two types of timing variables:

- Analysis timing variables
  - Describe the timing of the analysis value
  - Start with the letter "A"
  - Are directly associated with the AVAL and AVALC
- Supportive timing variables
  - Are not directly descriptive of the analysis value (AVAL and/or AVALC)
  - May be included to support review
  - Are often SDTM timing variables copied into analysis datasets to support data traceability and/or show how ADaM timing variables differ from SDTM data

Mapping of a BDS

BDS Variables – Indicator Variables

- Indicator variables are used to identify observations that are important to the analysis in some way
  - Indicator variables can be used for analysis or support
  - Describe each indicator variable in the metadata

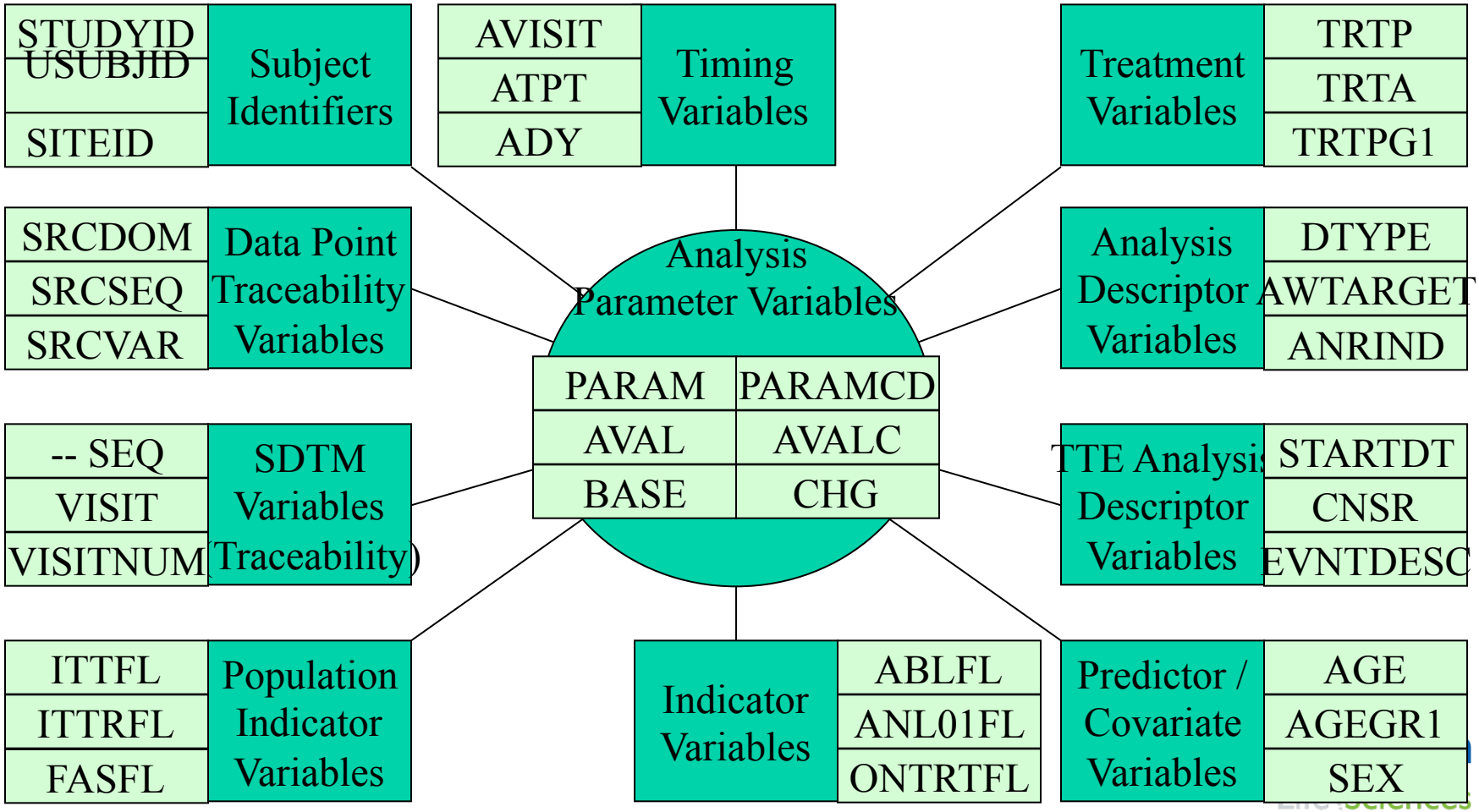
- Conventions:
  - Character variables end in *FL (Y/N/Null or Y/Null)
  - Numeric variables end in *FN (1/0/Null or 1/Null)
    - Can be included in addition to corresponding character variable

BDS Variables – Covariates in BDS

- Include covariates used for statistical analyses
  - Remember analysis-ready principle
- Some indicator variables may also serve as covariates for some statistical models

Current Standard ADaM Data Structures

Understanding the ADaM Basic Data Structure



Mapping of a BDS

Example ADEFF

Thank you

Brussels, Belgium, 18th September 2014

Business & Decision
Life Sciences

Jessica Minkue Mi Edou| Statistical Programmer| jessica.minkue@businessdecision.com

Business & Decision Life Sciences
Sint-Lambertusstraat 141 rue Saint-Lambert
B-1200 Brussels
T: +32 2 774 11 00
F: +32 2 774 11 99
lifesciences@businessdecision.com
http://www.businessdecision-lifesciences.com/