

Analysis Result Metadata... are we there yet?

Carla Santillan, Chiltern International bvba, Belgium



CHILTERN
Designed Around You®



PhUSE Annual Conference

BARCELONA 2016



Agenda

- Introduction
- Starting with Analysis Result Metadata
- Minimize Extra Work
- Creating Advantages
- Gain Speed and Efficiency
- Conclusions
- Questions

- Studies starting after March 15, 2018 must use Define 2.0
 - Analysis Result Metadata (ARM) is part of Define 2.0
 - Not mandatory until now
- Actual approach
 - ARM is manually done at the end of the TLF creation process
- Ambitious & efficient approach
 - Identify relationships between each element of ARM and its source, existent standard files
- Feasible? What can be done in our organization?

- Introduction
- **Starting with Analysis Result Metadata**
- Minimize Extra Work
- Creating Advantages
- Gain Speed and Efficiency
- Conclusions
- Questions

- ARM, sponsor decision, normally done for the primary and secondary efficacy results
- We need to establish traceability between a given analysis result and the specific source
 - ADaM dataset, Protocol, SAP, TLF specifications, SAS code
- At the moment if created, it's done at the end of the TLF creation process
 - Why?
 - Decision of submission, new standard, content not well known, process and roles are not defined
- Different programmer involved at the end for the creation of ARM
- Therefore the actual approach implies some **re work** at the moment of ARM creation, not only for each of the analysis results part of this metadata, but also for the next study to be submitted

Starting with Analysis Result Metadata

- ARM field's description and its source

Metadata Field	Description	Source
DISPLAY IDENTIFIER	An identifier for the specific analysis display (such as a display number), when combined with the result identifier provides a unique identifier for the analysis result.	Table/Figure number, in TLF metadata (*). Easy to identify
DISPLAY NAME	Title of display , including additional information that might be needed to describe and identify the display (e.g., analysis population).	Table/Figure name, in TLF metadata. Easy to identify
RESULT IDENTIFIER	Identifies specific analysis result when there are multiple analysis results on a display.	Table/Figure footnotes, in TLF metadata. Easy to identify
PARAM	The analysis parameter that is the focus of the analysis result.	From ADaM specifications. Easy to identify
PARAMCD	Corresponds to PARAM in the analysis dataset	From ADaM specifications. Easy to identify
ANALYSIS VARIABLE	The analysis variable being analyzed	From ADaM specifications. Easy to identify

Starting with Analysis Result Metadata

Metadata Field	Description	Source
REASON	The rationale for performing this analysis. It will indicate when the analysis was planned (e.g., 'Pre-specified in Protocol,' 'Pre-specified in SAP,' 'Data Driven,' 'Requested by FDA') and the purpose of the analysis within the body of evidence (e.g., 'Primary Efficacy', 'Key Secondary Efficacy', 'Safety').	Manual input from CP, from SAP or Protocol. Entails re-work, this information was clearly understood and available when the analysis result was programmed
DATASET	The name of the dataset(s) used to generate the analysis result	From ADaM specifications. Easy to identify
SELECTION CRITERIA	Specific and sufficient selection criteria for analysis subset and / or numerator – a complete list of the variables and their values used to identify the records selected for the analysis.	Manual input from CP, from SAP or Protocol. Entails re-work, this information was clearly understood and available when the analysis result was programmed
DOCUMENTATION	Textual description of the analysis performed.	Manual input from CP, from SAP or Protocol. Entails re-work, this information was clearly understood and available when the analysis result was programmed
PROGRAMMING STATEMENTS	The software code used to perform the specific analysis. This will include the model statement (using the specific variable names) and all technical specifications needed for reproducing the analysis (e.g. covariance structure).	From SAS ® code, manual input from CP. Entails re-work, this information was clearly understood and available when the analysis result was programmed

- (*) TLF metadata: Describes basic parameters per analysis output (table, listing, figure) within a standardize template, like table number, title, footnote, etc.

- How much re work is needed?
 - Assuming
 - Familiarization with ADaM, ARM content and purpose
 - Familiarization with the study
- Four out of eleven ARM elements needs re reading study specific standard documents
 - REASON, SELECTION CRITERIA, DOCUMENTATION, PROGRAMMING STATEMENTS
 - and four times multiplied per each analysis result included in ARM
- All those field's contents were natural questions at the moment of programming
- Not too difficult, but time consuming and we would try to avoid it

Starting with Analysis Result Metadata

- Can we minimize these extra work?
- Can we re use existing material? And standardize present tools?
- How do we gain speed and efficiency?

- Introduction
- Starting with Analysis Result Metadata
- **Minimize Extra Work**
- Creating Advantages
- Gain Speed and Efficiency
- Conclusions
- Questions

- Case study, standard files involved in the creation of ARM and its relationships

Program: </directory/directory/directory/program.sas>
Study: 99-NN-001
Table 12.3.1.1
Mean NRS Pain Score Over the Last 5 Days for Overall Pain. Inferential Analysis
Full Analysis Set

XXXX Version
Source: XXXX

Study Part: xxxxxxxx

Parameter	Statistic	Placebo (N=XXX)	DRUGA 3 mg (N=XXX)	DRUGA 5 mg (N=XXX)
EoT Change From Baseline	n	xxx	xxx	xxx
	Mean (SEM)	xx.xx (xx.xx)	xx.xx (xx.xx)	xx.xx (xx.xx)
	95% Two-sided CI	(xx.xx, xx.xx)	(xx.xx, xx.xx)	(xx.xx, xx.xx)
EoT Difference vs. Placebo	Mean (SEM) [2]		xx.xx (xx.xx)	xx.xx (xx.xx)
	95% Two-sided CI [2]		(xx.xx, xx.xx)	(xx.xx, xx.xx)
	p-value[1] p-value[2]	x.xxx	x.xxx	x.xxx
Linear Trend [3]	Estimate	x.xxxx		
	p-value	x.xxx		

SEM: Standard Error of the Mean. NE: Non-Estimable.
From ANCOVA model that includes treatment group and region as fixed factors and baseline value as a covariate.
[1] Overall treatment effect.
[2] From the pairwise comparisons of each active group vs. placebo based on Dunnett's test in the ANCOVA model.
[3] Overall treatment effect from linear trend.

Date ddMmmYYYY hh:mm:ss

Page X of Y

- The study measured daily pain scores. The data is mapped into SDTM.QS and then used in ADaM.ADQS. Where NRS, Numeric Rate Scale, questionnaire is used to give a numeric value to the intensity of pain
- As described in the protocol, one of the primary efficacy analysis variables is the percentage change from baseline to end of study (or early discontinuation) of the mean NRS pain score for overall pain.
- In the efficacy table chosen for this example, NRS scores were analysed using an ANCOVA model which included dose group and region (REG1 and REG2) as fixed factors and baseline NRS pain score of overall pain as a covariate.
- The analysis was conducted using the SAF population.

- ARM for the example

Metadata Field	Metadata
DISPLAY IDENTIFIER	Table 12.3.1.1
DISPLAY NAME	Mean NRS Pain Score Over the Last 5 Days for Overall Pain. Full Analysis Set
RESULT IDENTIFIER	Treatment difference results (Mean, confidence interval, p-value)
PARAM	Overall Pain Score during the 5-day Period
PARAMCD	PLPNOV
ANALYSIS VARIABLE	CHG, BASE, TRT02AN, GEOREGN
REASON	Primary efficacy analysis as pre-specified in protocol
DATASET	ADQS
SELECTION CRITERIA	fas1fl='Y', paramcd='PLPNOV', trt01pn~=. , avisit='EoT'
DOCUMENTATION	See Protocol Section XX for details. Program: program_ex1.sas. NRS scores were analysed using an ANCOVA model which included dose group and region (REG1 and REG2) as fixed factors and baseline NRS pain score of overall pain as a covariate.
PROGRAMMING STATEMENTS	<pre> data pain; set adam.adqs; where fas1fl='Y' and paramcd="PLPNOV" and avisit="EoT"; run; proc mixed data=pain; class &trt georegn; model chg=base &trt georegn; lsmeans &trt/cl adjust=dunnett; estimate 'Linear trend' &trt -2 -1 0 1 2; ods output type3=pvalue; ods output lsmeans=lsm; ods output diffs=dif; ods output estimates=trend; run; </pre>

Minimize Extra Work

- The figure below shows each element of ARM and its source standard file

Metadata Field	Metadata	
DISPLAY IDENTIFIER	Table 12.3.1.1	
DISPLAY NAME	Mean NRS Pain Score Over the Last 5 Days for Overall Pain. Full Analysis Set	TFL Metadata
RESULT IDENTIFIER	Treatment difference results (Mean, confidence interval, p-value)	
PARAM	Overall Pain Score during the 5-day Period	TFL specification
PARAMCD	PLPNOV	
ANALYSIS VARIABLE	CHG, BASE, TRT02AN, GEOREGN	
REASON	Primary efficacy analysis as pre-specified in protocol	
DATASET	ADQS	ADaM Metadata
SELECTION CRITERIA DOCUMENTATION	fas1fl='Y', paramcd='PLPNOV', trt01pn~=. , avisit='EoT' See Protocol Section XX for details. Program: program_ex1.sas. NRS scores were analysed using an ANCOVA model which included dose group and region (REG1 and REG2) as fixed factors and baseline NRS pain score of overall pain as a covariate.	
PROGRAMMING STATEMENTS	<pre>data pain; set adam.adqs; where fas1fl='Y' and paramcd="PLPNOV" and avisit="EoT"; run; proc mixed data=pain; class &trt georegn; model chg=base &trt georegn; lsmeans &trt/cl adjust=dunnett; estimate 'Linear trend' &trt -2 -1 0 1 2; ods output type3=pvalue; ods output lsmeans=lsm; ods output diffs=dif; ods output estimates=trend; run;</pre>	Protocol/SAP SAS code

- The information is coming from CDISC or company standard files
- Identify the elements needed to create ARM in the existing standards
- Identify relationships between standard files

Minimize Extra Work

- Standard file Available: ADaM specifications

of the dataset, used for	Qualifying Variable	Variable Name	Variable Label	ble Type (character)	Origin of the Variable	cdisc rol	Variable Length	Variable Codelist	Variable Comment
ADQS	PARAMCD	BPISMEAN	BPI Pain Severeance Total Score	Num	Derived		8		PARCAT1='BPI SHORT FORM', PARCAT2='BPI PAIN SEVERITY',

Domain Name	Variable Name	Variable Label	Type	Origin	Length	Codelist	Comment
ADQS	CHG	Change from Baseline	Num	Derived	8		AVAL-BASE, assign only on post-baseline records. Leave EMPTY if either AVALC or BASEC is equal to 'NOT APPLICABLE'.
ADQS	CHGCAT1	Change from Baseline Category 1	Char	Derived	200		If CHG=0 then assign 'NO CHANGE'. Else if CHG > 0 then assign 'INCREASE'. Else if . < CHG < 0 and CHG is not missing then assign 'DECREASE'.
ADQS	CHGCAT1N	Change from Baseline Category 1 (N)	Num	Derived	8	CHGCAT	Coded from CHGCAT1 (See Codelist).
ADQSAH	PCHG	Percent Change from Baseline	Num	Derived	8		100*(AVAL-BASE)/BASE, assign only on post-baseline records. Leave EMPTY if either AVALC or BASEC is equal to 'NOT APPLICABLE' or if BASE=0.

Minimize Extra Work

- Standard file Available: TLF Metadata

Program name	Output Name	Output Type	Output Number	Output Title	Title 1	Footnote 2	Footnote 3	Footnote 9	Source	Program Location	Developer	Verifier	Comments
t_121121_as	_2	Table	12.1.1.2.2.1	Analysis Sets in REG1		[1] All randomized subjects who took at least one dose of study medication (DRUGA or DRUGB or placebo).	[2] All randomized subjects who received at least one dose of double-blind study medication and who have a primary NRS pain score for overall pain		ADSL	directory/directory/directory/program_name.sas	CPA	CPB	
example1	_1	Table	12.3.1.1	Mean NRS Pain Score Over the Last 5 Days for Overall Pain. Inferential Analysis		Includes all treatment groups and information until Visit 6 (included).	Last non-missing observation before first intake of study drug is used as baseline.		ADQSS	directory/directory/directory/program_name.sas	CPA	CPB	
example1	_2	Table	12.3.1.2	Mean NRS Pain Score Over the Last 5 Days for Overall Pain. Inferential Analysis		Only includes subjects randomized to placebo in Part 1 and taking DRUGA in Part 2. Columns refer to DRUGA treatment from Visit 6 onwards.	Last non-missing observation before first dose of DRUGA is used as baseline.		ADQSAH	directory/directory/directory/program_name.sas	CPA	CPB	

- Sponsor Standard file Available: SAP

7.4	Analysis of Efficacy	19
7.4.1	Analysis of Primary Endpoint(s)	19
7.4.2	Analysis of Secondary Endpoints	22
7.4.3	Analysis of Exploratory Endpoints	22
7.4.4	Analysis of Other Variables	22
7.5	Analysis of Safety	22
7.5.1	Adverse Events	22
7.5.2	Clinical Laboratory Evaluation	25
7.5.3	Vital Signs	28
6.5	Other Variables	15
7	STATISTICAL METHODOLOGY	15
7.1	General Considerations	15

- Sponsor Standard file Available: SAP; TLF Specifications

IV.	STANDARD TLFs BY DOMAIN	13
A.	Disposition	13
1.	Guidelines for Disposition TLFs	13
2.	Tables	15
3.	Listings	22
B.	Demographics	33
1.	Guidelines for Demographic TLFs	33
2.	Tables	34
3.	Listings	38
C.	Medical History	39
1.	Guidelines for Medical History TLFs	39
2.	Tables	40
3.	Listings	42

- Introduction
- Starting with Analysis Result Metadata
- Minimize Extra Work
- **Creating Advantages**
- Gain Speed and Efficiency
- Conclusions
- Questions

- Identify the elements of interest in each standard file, can it be modified to facilitate the creation of ARM?
- We cannot easily use the contents from a Word file (Protocol, SAP, TLF specs.).
- Some possible improvements
 - Vocabulary used in interrelated standards documents should be consistent
 - Training including practical examples of Analysis Result Metadata should be available for statisticians and clinical programmers
- TLF metadata can be easily improve

Creating Advantages: TLF metadata improvements

Program name	Output Name	Output Type	Output Number	Title 3	Title 10	Footnote 2	Footnote 3	Footnote 9	Source	ARM	Program Location	Developer	Verifier	Comments
t_121121_a	_2	Table	1	Analysis Sets in REG1		[1] All randomized subjects who took at least one dose of study medication (DRUGA or DRUGB or placebo).	[2] All randomized subjects who received at least one dose of double-blind study medication and who have a primary NRS pain score for overall pain		ADSL	N	directory/directory/directory/program_name.sas	CPA	CPB	
example1	_1	Table	1.1	Mean NRS Pain Score Over the Last 5 Days for Overall Pain. Inferential Analysis		Includes all treatment groups and information until Visit 6 (included).	Last non-missing observation before first intake of study drug is used as baseline.		ADQS	Y	directory/directory/directory/program_name.sas	CPA	CPB	
example1	_2	Table	12.3.1.2	Mean NRS Pain Score Over the Last 5 Days for Overall Pain. Inferential Analysis		Only includes subjects randomized to placebo in Part 1 and taking DRUGA in Part 2. Columns refer to DRUGA treatment from Visit 6 onwards.	Last non-missing observation before first dose of DRUGA is used as baseline.		ADQSAH	Y	directory/directory/directory/program_name.sas	CPA	CPB	
example1	_3	Table	12.3.1.3	Mean NRS Pain Score Over the Last 5 Days for Overall Pain. Inferential Analysis		Includes subjects within the same treatment group during the whole study and information for all visits.	Last non-missing observation before first intake of study drug is used as baseline.		ADQSAH	Y	directory/directory/directory/program_name.sas	CPA	CPB	

- Extra tab to include the link between Analysis result and the output (table, or figure).
- From Protocol and SAP, the list of TLFs included in ARM should be identified, and added into TLF metadata
- Content of ARM can be QC-ed by the clinical programmer who is in charge of dual programming

Creating Advantages: TLF metadata improvements

Program name	Output Name	DISPLAY IDENTIFIER	Developer	Verifier	DATASET	PARAMCD (CP)	ANALYSIS VARIABLE (CP)	RESULT IDENTIFIER (STAT)	REASON (STAT)	DOCUMENTATION (STAT)	SELECTION CRITERIA (CP)	PROGRAMMING STATEMENTS (CP)	Status
example 1	1	Table 12.3.1.1	CPA	CPB	ADQS	PLPNOV	PCHG	Treatment difference results (Mean, confidence interval, p-value)	Primary efficacy analysis as pre-specified in protocol	See Protocol Section XX for details. Program: program_ex1.sas. NRS scores were analysed using an ANCOVA model which included dose group and region (REG1 and REG2) as fixed factors ...	fas1fl='Y', param='PLPNOV', trt01pn~., avis='EoT'	proc mixed data=pain; class &trt georegn; model chg=base &trt georegn; lsmeans &trt/cl...	Not done Passed QC QC findings
example 1	1	Table 12.3.1.1	CPA	CPB	ADQS			Pairwise comparison...					
example 1	2	Table 12.3.1.2	CPA	CPB	ADQS								
example 1	3	Table 12.3.1.3	CPA	CPB	ADQS								

Navigation: TFL Metadata | **ARM** | +

- Extra tab: use information that uniquely characterise an output and add ARM elements to be filled by CP=Clinical programmer or STAT= Statistician

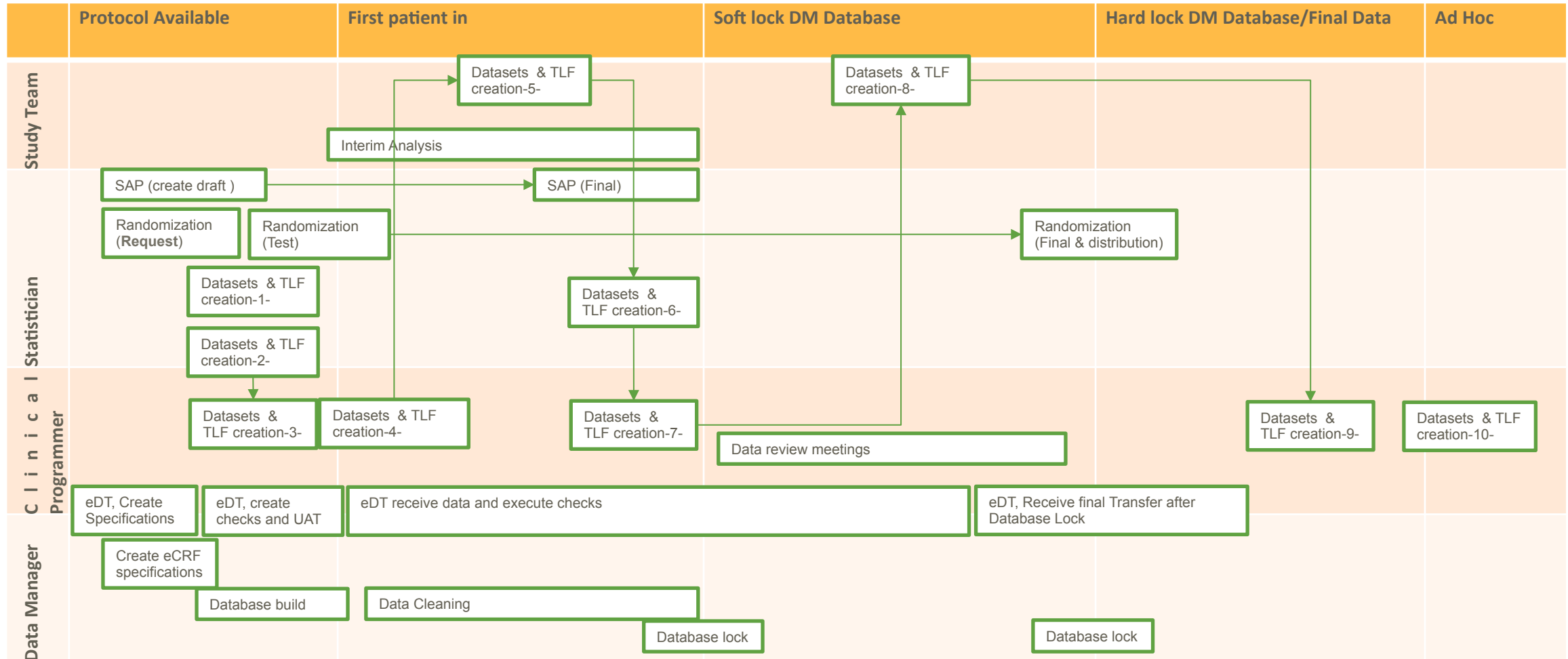
- Introduction
- Starting with Analysis Result Metadata
- Minimize Extra Work
- Creating Advantages
- **Gain Speed and Efficiency**
- Conclusions
- Questions

Gain Speed and Efficiency



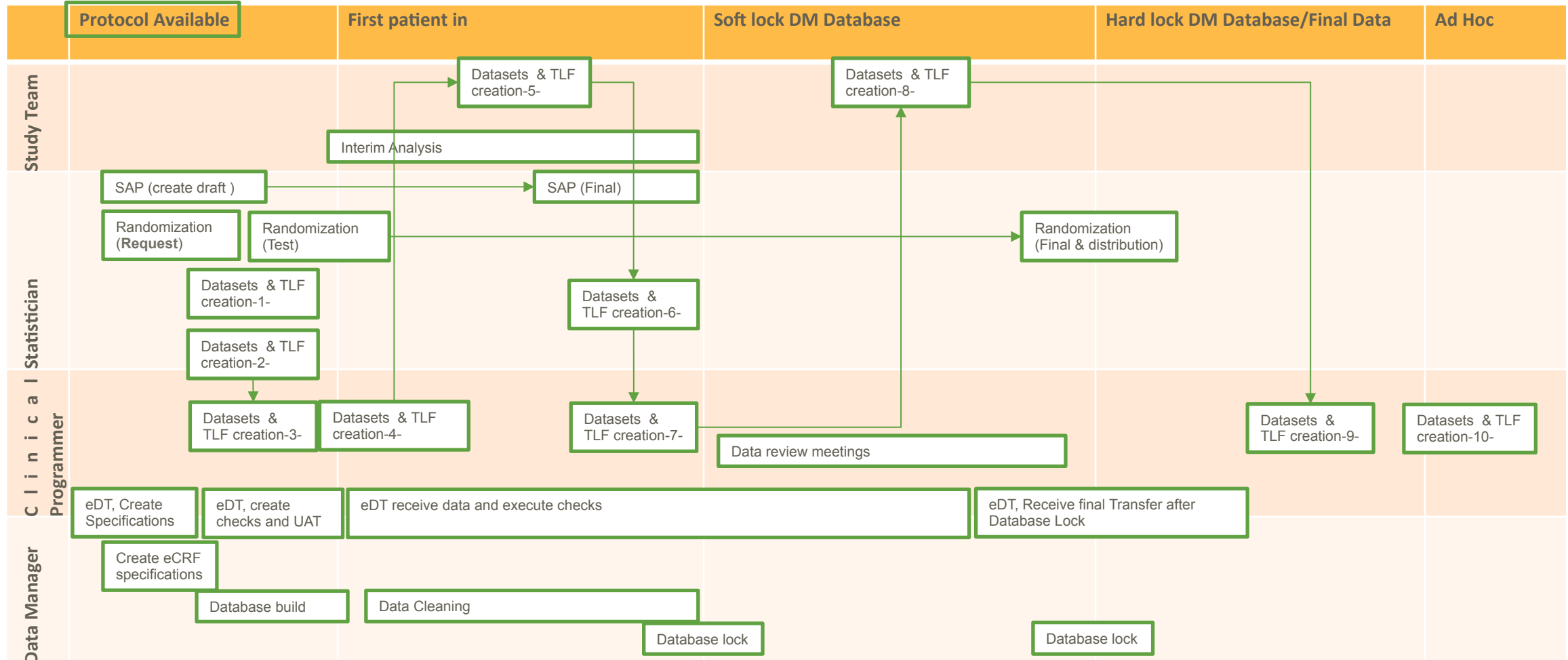
- Inputting the correct information once and at the most opportune time
- Which relationships we can create between standard files and when in the process of TLF creation

Gain Speed and Efficiency



- Note that the example above is a simplified version of the process that our sponsor follows, it may be different in other companies

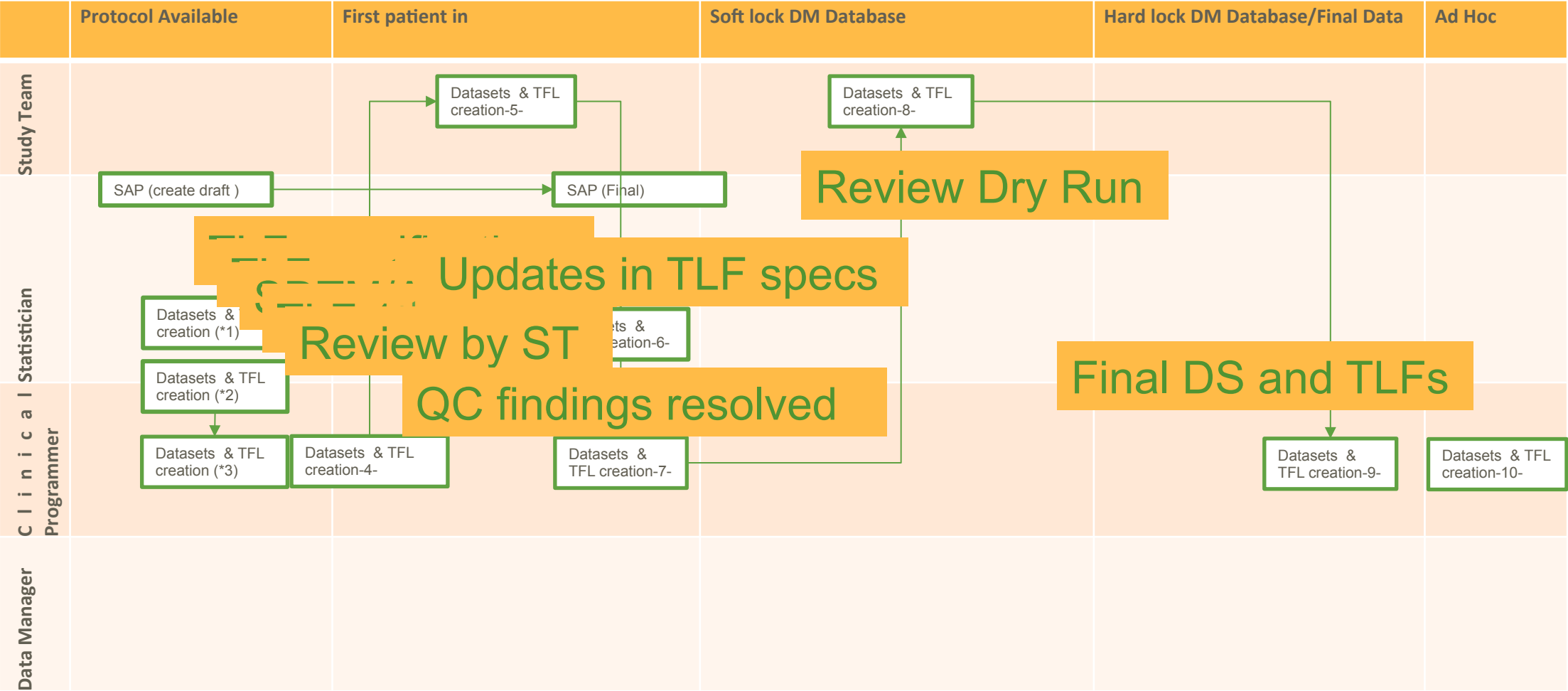
Gain Speed and Efficiency



- Note that the example above is a simplified version of the process that our sponsor follows, it may be different in other companies

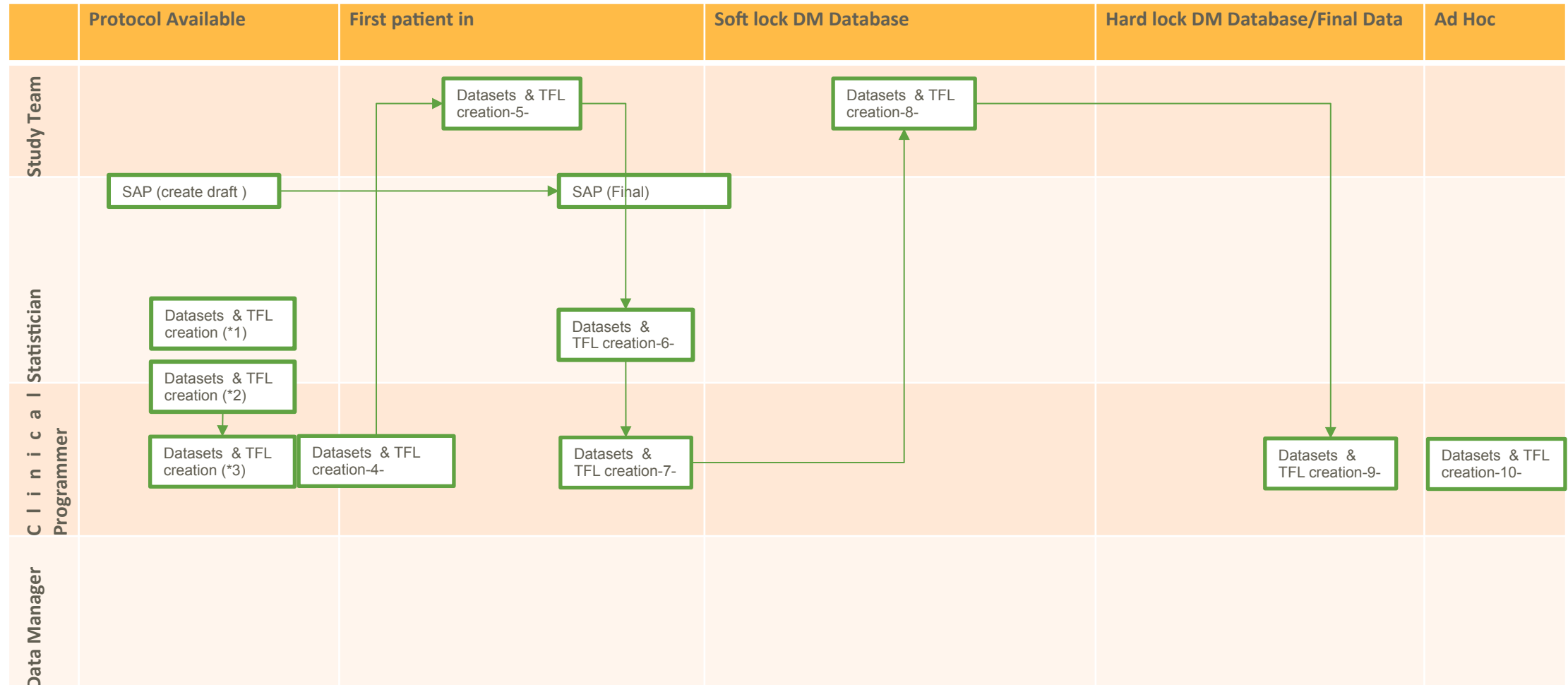
Gain Speed and Efficiency

- When and by who?



Gain Speed and Efficiency

- Best point in the process to start creating ARM is at the time of TLF metadata creation



Gain Speed and Efficiency

Program name	Output Name	DISPLAY IDENTIFIER	Developer	Verifier	DATASET	PARAMCD (CP)	ANALYSIS VARIABLE (CP)	RESULT IDENTIFIER (STAT)	REASON (STAT)	DOCUMENTATION (STAT)	SELECTION CRITERIA (CP)	PROGRAMMING STATEMENTS (CP)	Status
example 1	1	Table 12.3.1.1	CPA	CPB	ADQS	PLPNOV	PCHG	Treatment difference results (Mean, confidence interval, p-value)	Primary efficacy analysis as pre-specified in protocol	See Protocol Section XX for details. Program: program_ex1.sas. NRS scores were analysed using an ANCOVA model which included dose group and region (REG1 and REG2) as fixed factors ...	fas1fl='Y', param='PLPNOV', trt01pn~., avis='EoT'	proc mixed data=pain; class &trt georegn; model chg=base &trt georegn; lsmeans &trt/cl...	Not done Passed QC QC findings
example 1	1	Table 12.3.1.1	CPA	CPB	ADQS			Pairwise comparison...					
example 1	2	Table 12.3.1.2	CPA	CPB	ADQS								
example 1	3	Table 12.3.1.3	CPA	CPB	ADQS								

TFL Metadata **ARM** +

- Now we can complete who can create the element and specify when ARM can be created

- Introduction
- Starting with Analysis Result Metadata
- Minimize Extra Work
- Creating Advantages
- Gain Speed and Efficiency
- **Conclusions**
- Questions

- Within the organization, possible relationships between existent standard files that could facilitate the creation of ARM were identified
- In each standard source file for ARM the following was looked at:
 - Who is the author and when is the standard file created
 - How can standard files be related
 - Possible additions in existent standard files
- ARM training, with practical examples.
- A review of the complete process, what is created when and by whom can help to identify who would be most suitable during the different stages of a study for the completion of ARM and for the reviewing/QCing of ARM

- We should think in ARM as an internal needed information to have a good understanding of the critical tables instead of taking it as an obligatory submission document that will be needed for studies starting in March 15, 2018!

Questions

