

innovation

Challenges of using ETL tools to deliver infinitely flexible clinical data

May 20, 2015

About the Presenter

- Over 18 Years SAS Experience
 - PAREXEL
 - Takeda
 - Lilly
 - Dootsonic
 - d-Wise
 - + others...
- PhUSE Committee member
- Coding Solutions Stream Chair

About the Presenter



About the Presenter



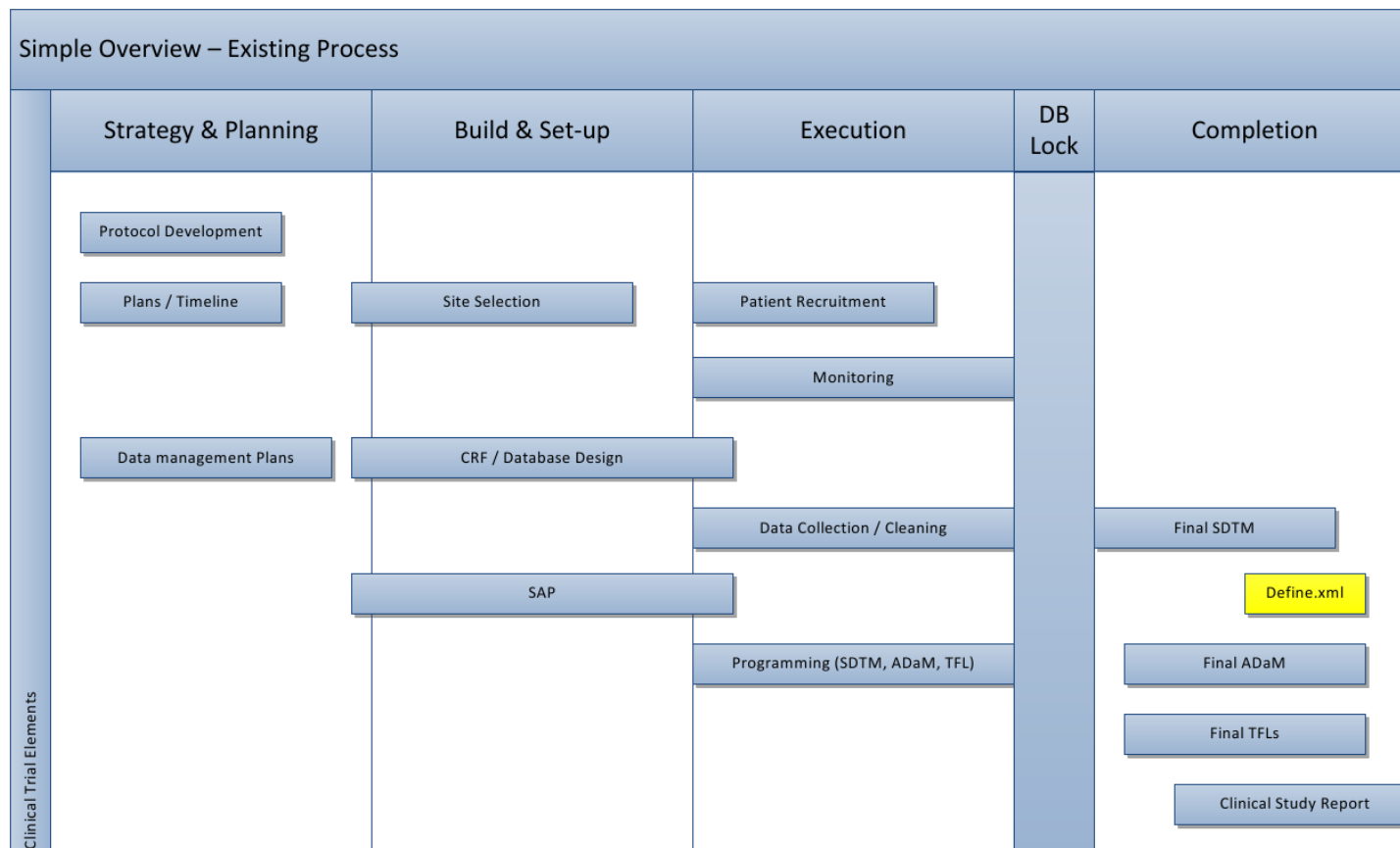
End to End Process

- What is driving your process?
 - Is it the science?
 - Or is it the standards?

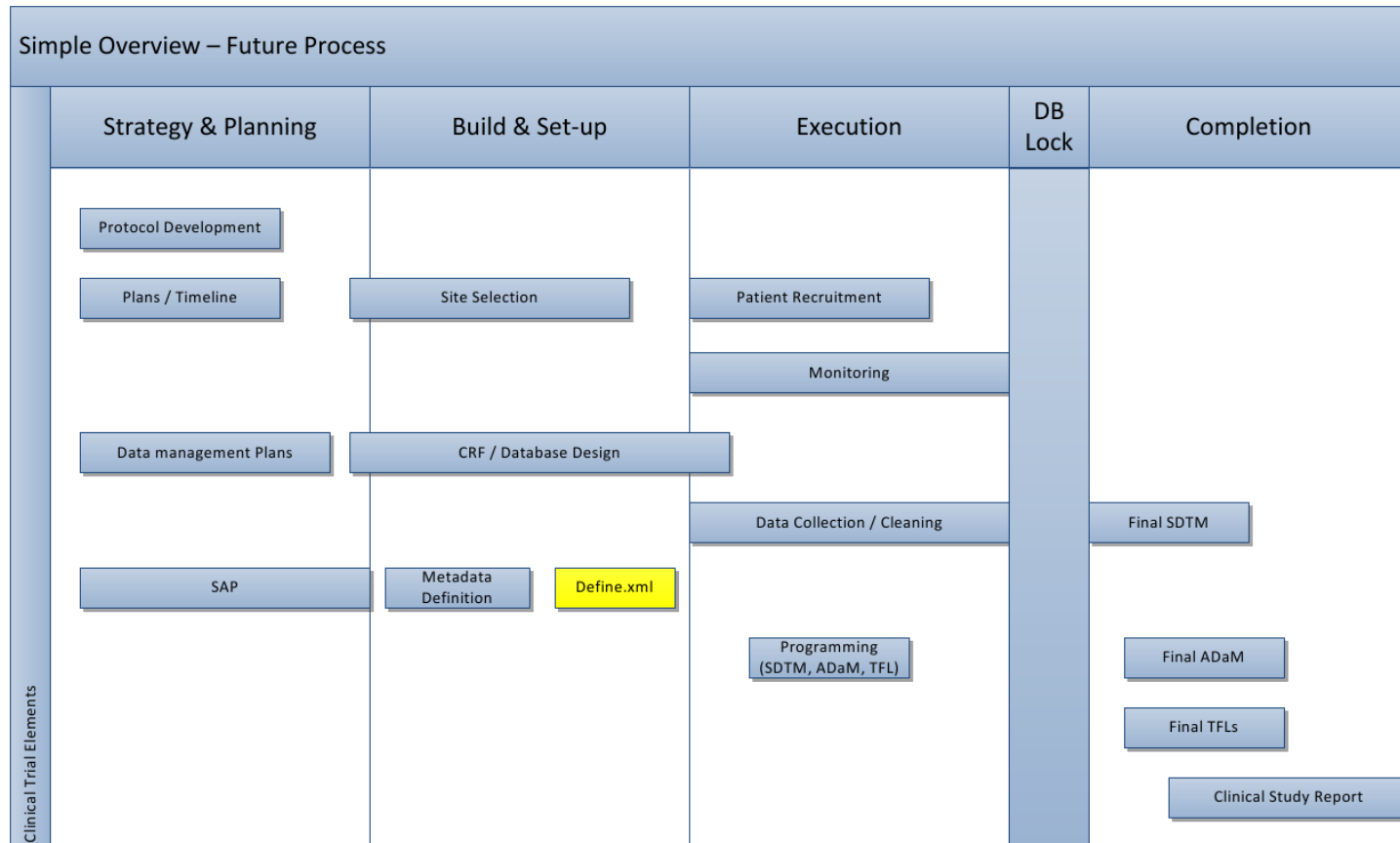
The Process

- Study Design
- Data Collection
- Data Management
- Statistical Programming
- Statistics
- Reporting

The Process



The Process



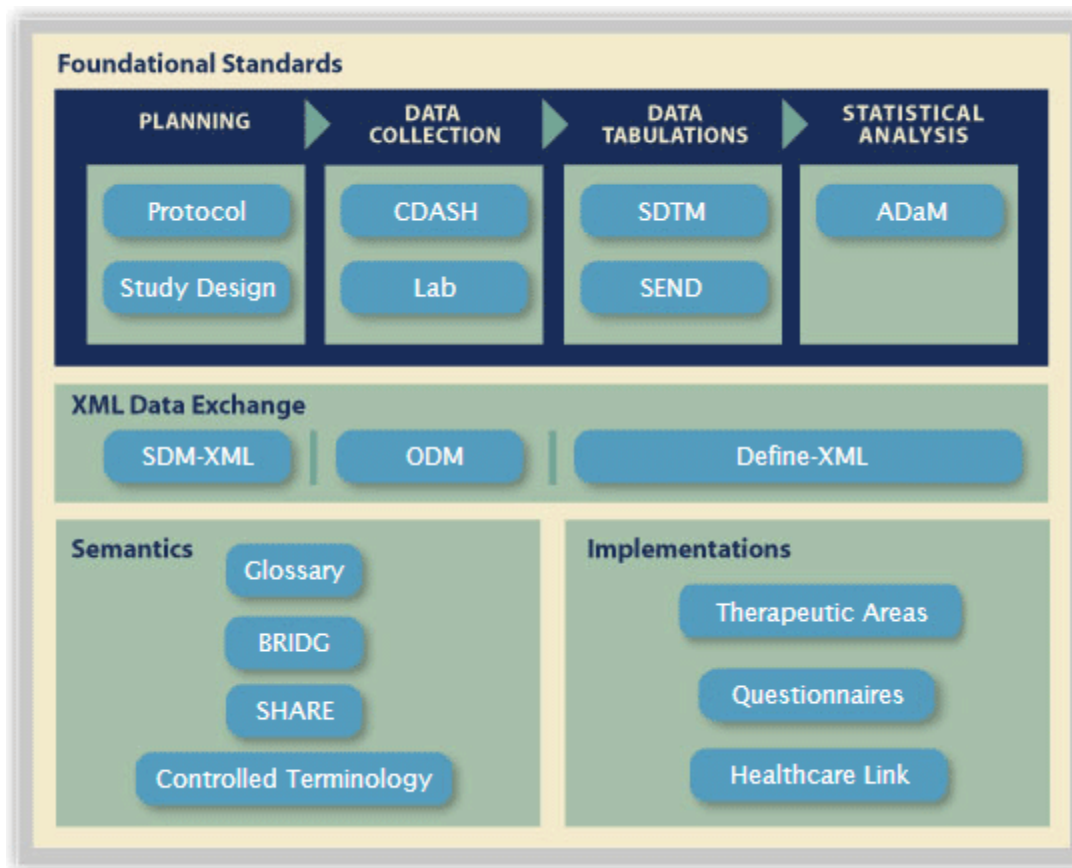
Science vs. Standards



phillipmartin.info

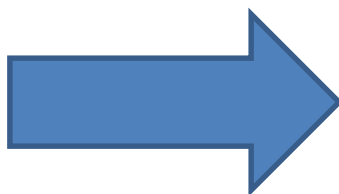


Science vs. Standards



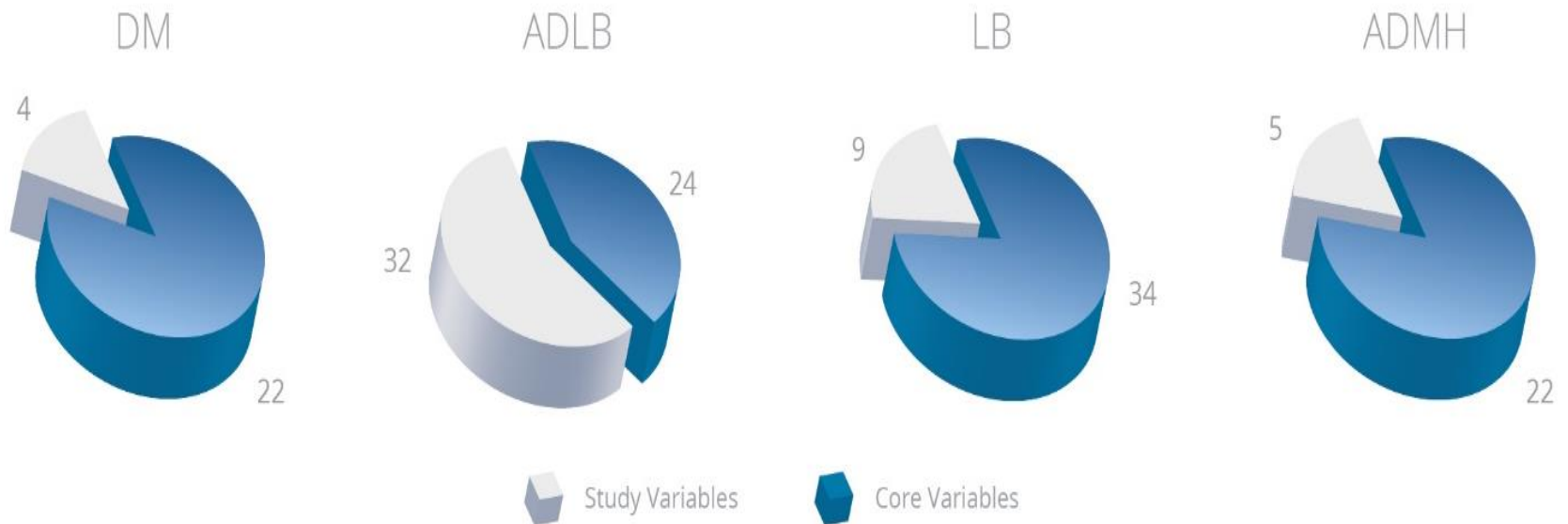
Standard Programming

STUDYID	INVID
DOMAIN	INVNAM
USUBJID	BRTHDTC
SUBJID	AGE
RFSTDTC	AGEU
RFENDTC	SEX
RFXSTDTC	RACE
RFXENDTC	ETHNIC
RFICDTC	ARMCD
RFPENDTC	ARM
DTHDTC	ACTARMCD
DTHFL	ACTARM
SITEID	COUNTRY



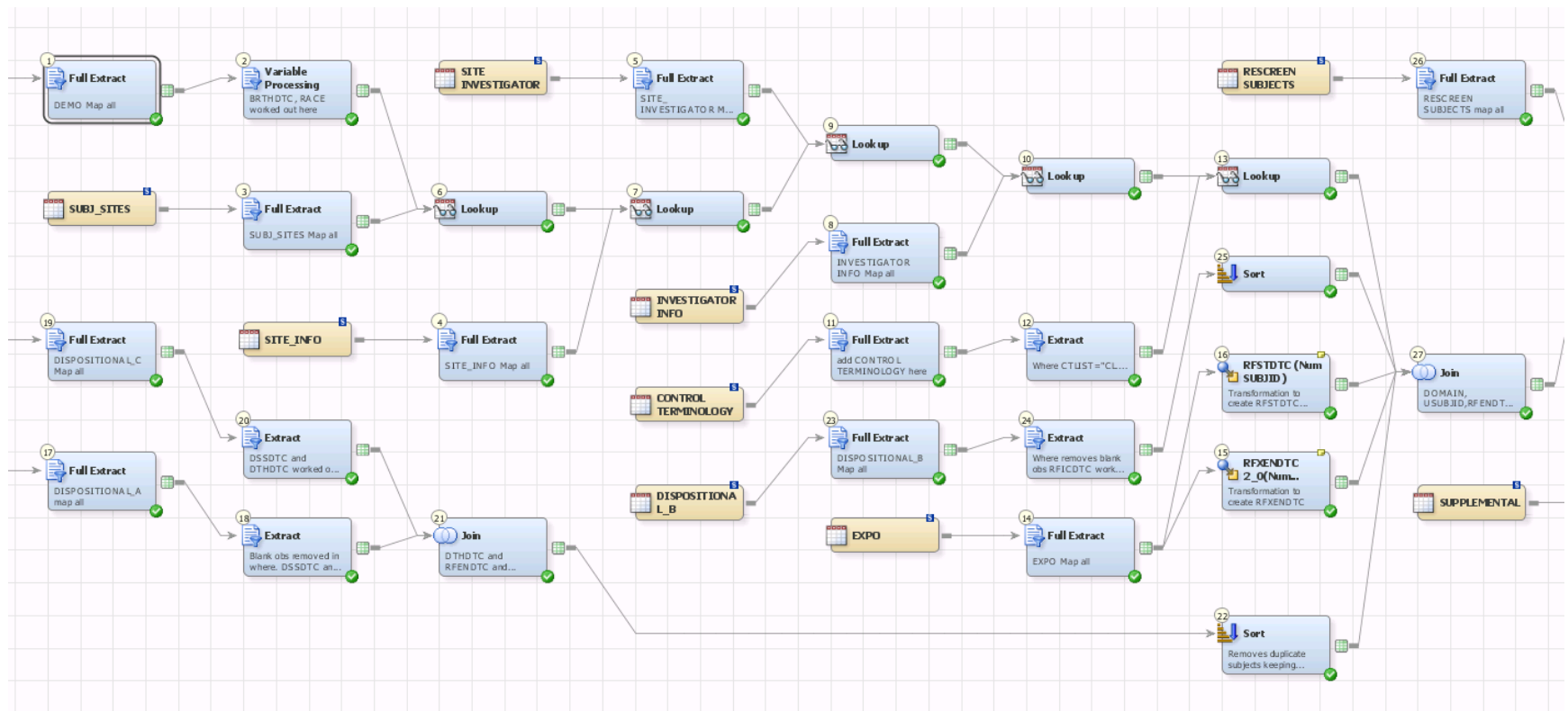
STUDYID	INVID
DOMAIN	INVNAM
USUBJID	BRTHDTC
SUBJID	AGE
RFSTDTC	AGEU
RFENDTC	SEX
RFXSTDTC	RACE
RFXENDTC	ETHNIC
RFICDTC	ARMCD
RFPENDTC	ARM
DTHDTC	ACTARMCD
DTHFL	ACTARM
SITEID	COUNTRY

Core vs. Study

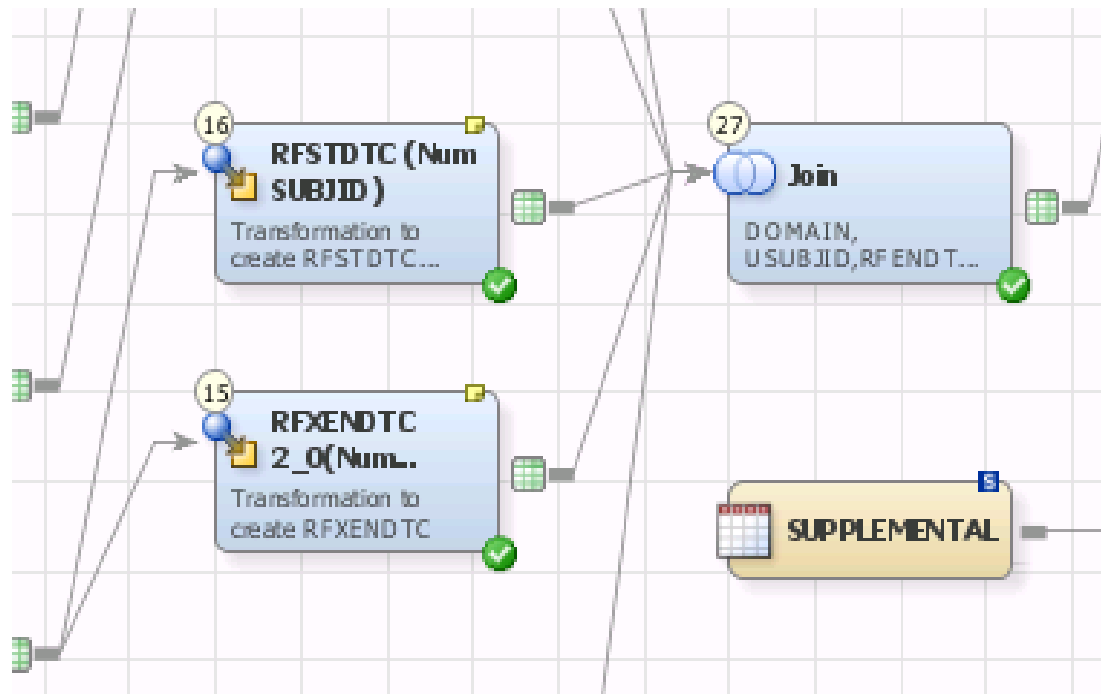


Blue represents variables that no longer need to be programmed. Grey represents variables that will need to be coded and validated

ETL



ETL



ETL

Table Loader Properties

General | Load Technique | Mappings | Options | Table Options | Code | Precode and Postcode | Status Handling | Parameters | Notes | Ext

Source table: Sort (dmsorted) | Target table: DM (DM)

#	Column	Column Description	#	Column	Column Description
1	STUDYID		1	STUDYID	Study Identifier
2	DOMAIN		2	DOMAIN	Domain Abbreviation
3	USUBJID		3	USUBJID	Unique Subject Identifier
4	SUBJID		4	SUBJID	Subject Identifier for the Study
5	RFSTDTC		5	RFSTDTC	Subject Reference Start Date/Time
6	RFENDTC		6	RFENDTC	Subject Reference End Date/Time
7	RFXSTDTC		7	RFXSTDTC	Date/Time of First Study Treatment
8	RFXENDTC		8	RFXENDTC	Date/Time of Last Study Treatment
9	RFICDTC		9	RFICDTC	Date/Time of Informed Consent
10	RFPENDTC		10	RFPENDTC	Date/Time of End of Participation
11	DTHDTC		11	DTHDTC	Date/Time of Death
12	DTHFI		12	DTHFI	Subject Death Flag

The Problem

- The problem is the dream we have been sold of standards, of tools that can ease our pain by automating the whole process.
- The standards are a good start, the tools are effective. But you cannot have infinite flexibility around the science coupled with the rigidity of ETL tools and expect it to be smooth sailing.

The Solution

- Metadata Driven Systems
- RDF / Direct Graphs
- Evolution of Standards
- Flexibility and management of standards