

ALSC

AUTOMATED CREATION OF SUBMISSION-READY ARTIFACTS

SILAS MCKEE

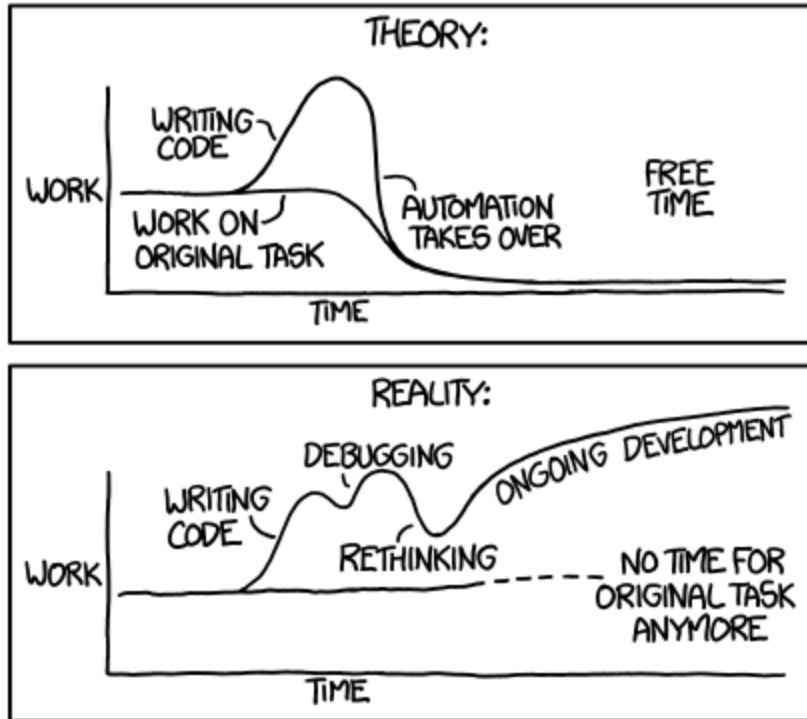


AGENDA

- 1. Motivation**
- 2. Automation Overview**
- 3. Architecture**
- 4. Validating the System**
- 5. Pilot Study Results**
- 6. Future State**

WHY AUTOMATE?

"I SPEND A LOT OF TIME ON THIS TASK.
I SHOULD WRITE A PROGRAM AUTOMATING IT!"



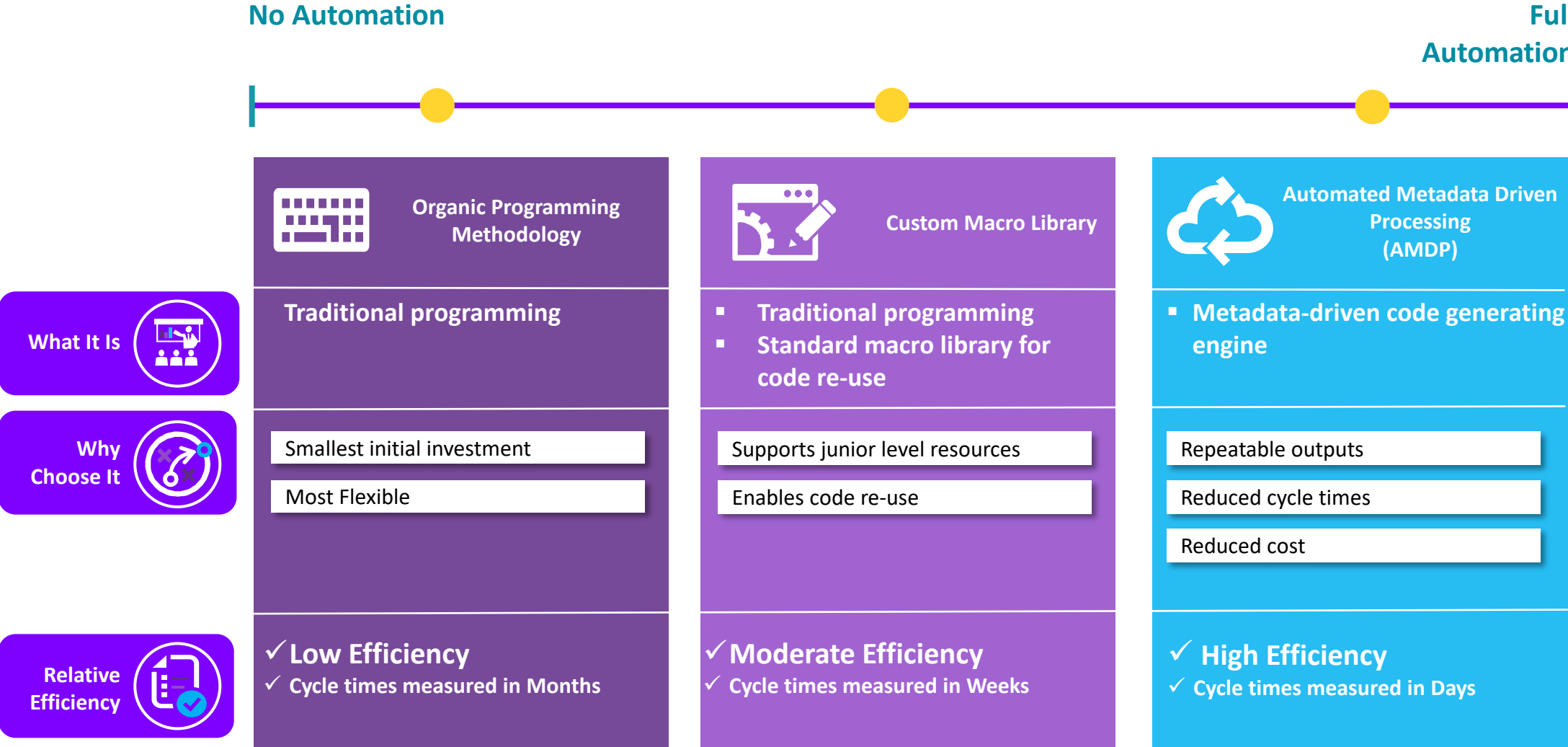
Source: <https://xkcd.com/1319/>

$$\frac{\text{Return (benefit)}}{\text{Investment (cost)}} = ROI \text{ (Return on Investment)}$$

- Reduce the per trial cost of submission artifact creation
- Reduce the cycle time for submission artifact creation
- Increase data quality

APPROACHES TO SUBMISSION ARTIFACT CREATION

Moving from traditional programming to full automation



AUTOMATION WITH METADATA

CDISC Metadata

Vital Signs Dataset (VS)						
Variable	Label	Type	Controlled Terms or Format	Origin	Role	Comment
STUDYID	STUDY IDENTIFIER	text		CRF Page	Identifier	Demographics CRF Page 4
DOMAIN	DOMAIN ABBREVIATION	text		CRF Page	Identifier	DOMAIN ABBREVIATION
USUBJID	UNIQUE SUBJECT IDENTIFIER	text		CRF Page	Identifier	Demographics CRF Page 4
VSSEQ	SEQUENCE NUMBER	integer		CRF Page	Identifier	SEQUENCE NUMBER
VSTESTCD	VITAL SIGNS TEST SHORT NAME	text		CRF Page	Topic	Vital Signs CRF Page 4, CRF Page 7
VSTEST	VITAL SIGNS TEST NAME	text		CRF Page	Synonym Qualifier	VITALS SIGNS TEST NAME
VSPOS	VITAL SIGNS POSITION OF SUBJECT	text		CRF Page	Record Qualifier	POSITION OF THE SUBJECT DURING A MEASUREMENT OR EXAMINATION.
VSORRES	RESULT OR FINDING IN ORIGINAL UNITS	text		CRF Page	Result Qualifier	Vital Signs CRF Page 4, CRF Page 7
VSORRESU	ORIGINAL UNITS	text		CRF Page	Variable Qualifier	ORIGINAL UNITS

Additional 'Operational' Metadata

SOURCE_DATASET_NAME	SOURCE_VAR_NAME	RULE_TYPE	DERIVATION_TYPE	DERIVATION_ORDER	RULE_DEFINITION	TARGET_DATASET_NAME	TARGET_VAR_NAME
VS_INPUT_TMP2	VSPULSE_STD VSPULSE_RAW	S	SC		75	VS_PULSE	VSSTRESN
VS_INPUT_TMP2	VSSTRESN	S	SC		76	VS_PULSE	VSSTRESC
VS_INPUT_TMP2	VSPULSE_STD_UN	D	AS		77 'beats/min'	VS_PULSE	VSSTRESU
VS_INPUT_TMP2	VSPULSE_STD	S			78	VS_PULSE	VSPOS
VS_INPUT_TMP2	VSMETHOD_PULSE_STD	S			79	VS_PULSE	VSMETHOD
VS_INPUT_TMP2(WHERE=(VSHR_VSHR_RAW	VSHR_RAW	D	AS		80 'HR'	VS_HR	VSTESTCD
VS_INPUT_TMP2	VSHR_RAW	D	AS		81 'Heart Rate'	VS_HR	VSTEST
VS_INPUT_TMP2	VSHR_RAW	S			82	VS_HR	VSORRES
VS_INPUT_TMP2	VSHR_UN	D	AS		83 'beats/min'	VS_HR	VSORRESU
VS_INPUT_TMP2	VSHR_STD VSHR_RAW	S	SC		84	VS_HR	VSSTRESN
VS_INPUT_TMP2	VSSTRESN	S	SC		85	VS_HR	VSSTRESC
VS_INPUT_TMP2	VSHR_STD_UN	D	AS		86 'beats/min'	VS_HR	VSSTRESU
VS_INPUT_TMP2	VSPULSE_STD	S			87	VS_HR	VSPOS
VS_INPUT_TMP2	VSMETHOD_HR_STD	S			88	VS_HR	VSMETHOD
VS_INPUT_TMP2(WHERE=(VBSA_VBSA_RAW	VBSA_RAW	D	AS		89 'BSA'	VS_BSA	VSTESTCD
VS_INPUT_TMP2	VBSA_RAW	D	AS		90 'Body Surface Area'	VS_BSA	VSTEST
VS_INPUT_TMP2	VBSA_RAW	S			91	VS_BSA	VSORRES
VS_INPUT_TMP2	VBSA_UN	S			92	VS_BSA	VSORRESU
VS_INPUT_TMP2	VBSA_STD VBSA_RAW	S	SC		93	VS_BSA	VSSTRESN
VS_INPUT_TMP2	VSSTRESN	S	SC		94	VS_BSA	VSSTRESC
VS_INPUT_TMP2	VBSA_STD_UN VBSA_UN	S	SC		95	VS_BSA	VSSTRESU
VS_INPUT_TMP2(WHERE=(VSHD_VSHDCIRC_RAW	VSHDCIRC_RAW	D	AS		96 'HDCIRC'	VS_HDCIRC	VSTESTCD
VS_INPUT_TMP2	VSHDCIRC_RAW	D	AS		97 'Head Circumference'	VS_HDCIRC	VSTEST

1. CDISC Metadata

- Provides structural metadata to inform dataset structure and variable creation statements
- Ensures compliance with data standards

2. Operational Metadata

- Provides instructions for join / merge operations
- Provides instructions for macro calls including parameters values
- Provides instructions on the algorithms used to transform, derive, or compute data.

METADATA MODEL

Complexity vs. Expressiveness

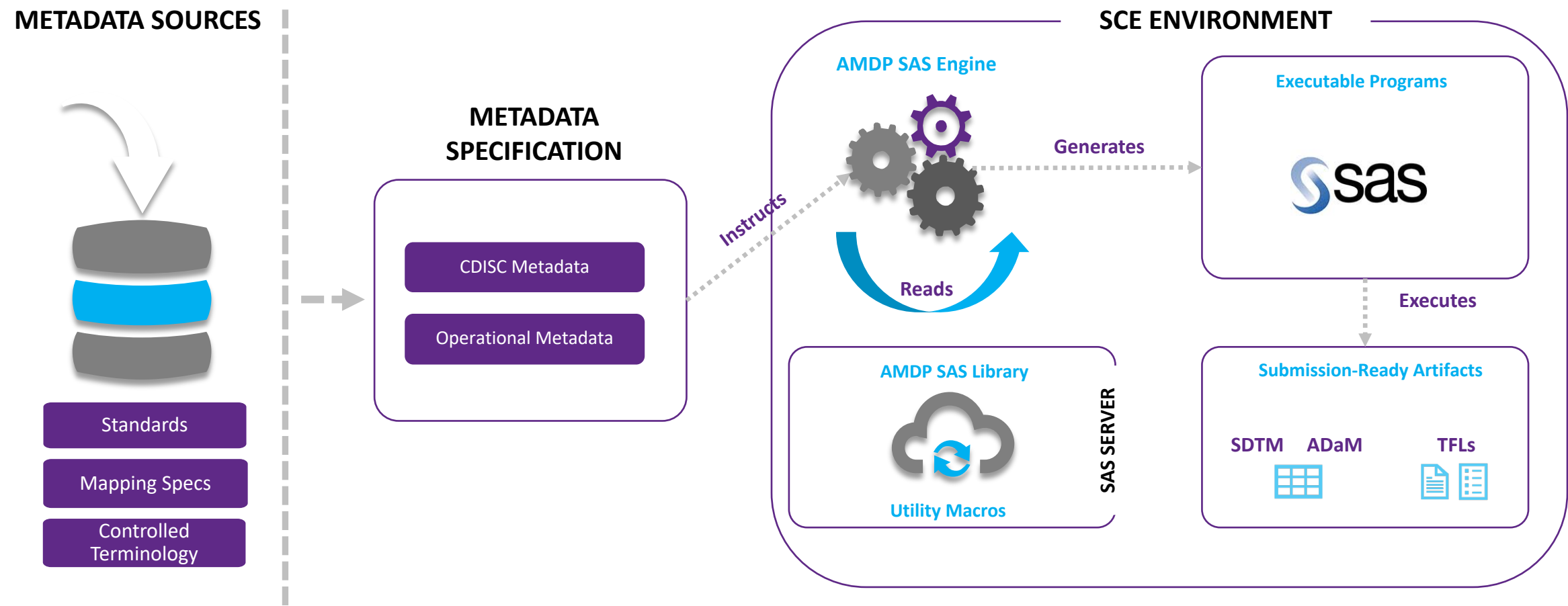
There is a natural trade-off between the complexity of the metadata model and its expressiveness.

	Study Metadata	Table Metadata	Metadata Complexity	Expressiveness
1	NA	Title: 'Summary Table for Vital Signs'	Simple	Low
2	Alignment: Left Font: Times New Roman Text Size: 12	Title: 'Summary Table for Vital Signs'	Medium	Medium
3	NA	Title: 'Summary Table for Vital Signs' Alignment: Left Font: Times New Roman Text Size: 12	Complex	High

Trivial Example

AMDP ARCHITECTURE

Current State



VALIDATING THE SYSTEM

The Artifacts

1. Created a set of standard artifacts that included representation from the following categories across SDTM, ADaM, and TFLs

SDTM

General Observation Classes

- Findings
- Interventions
- Events

Special Purpose

Special Purpose Relationship

ADaM

ADaM Subject-Level Analysis Dataset (ADSL)

Occurrence Data (OCCDS)

Basic Data Structure (BDS)

TFL

Basic Listings

Summary Tables

Change from Baseline Tables

Shift Tables

2. Validated the generation of these standard artifacts against data from two different studies

VALIDATING THE SYSTEM

Artifact Details

Artifact Type	Template or Class	Artifact(s) Produced during Validation
SDTM	Findings	DM, DS, EX, EG, VS, LB, IE, PE,
	Interventions	CM,
	Events	AE, HO
	Special Purpose	DM
ADaM	ADSL	ADSL
	OCCDS	ADCM
	BDS	ADEG
TFL	Summary Table – Continuous and Categorical Variables Statistics	“Demography Summary Table”
	Summary table – Frequency Count and Percent by Single or Double Level	“Concomitant Medications [by CMDECOD]”
	Summary table – Frequency Count and Percent by Single or Double Level	“Concomitant Medications within Classes of Interest [CMCLAS*CMDECOD]”
	Change from Baseline Table	“Summary of ECG Parameters and Change from Baseline by Time by Treatment”
	Shift Table	“Summary of Post-Baseline Worst-Case ECG Parameters Shift from Baseline by Treatment”
	Basic Listing	“Listing of Concomitant Medications”

PILOT STUDY RESULTS

Description

- The team received a proprietary global metadata library and source data for the study
- The team then developed study specifications (both human and machine-readable) to define the conversion of the source data into 48 SDTM domains.
- Successfully completed complete SDTM conversion with high quality in 31 business days.

Quality Metrics

Error Category	Error Count	Total Variables	Quality
Major	3	700	99.6%
Minor	6	700	99.1%
Total	9	700	98.7%

$$Quality = 1 - \left(\frac{Error\ Count}{Total\ Variables} \right)$$

PILOT STUDY RESULTS

Work Breakdown

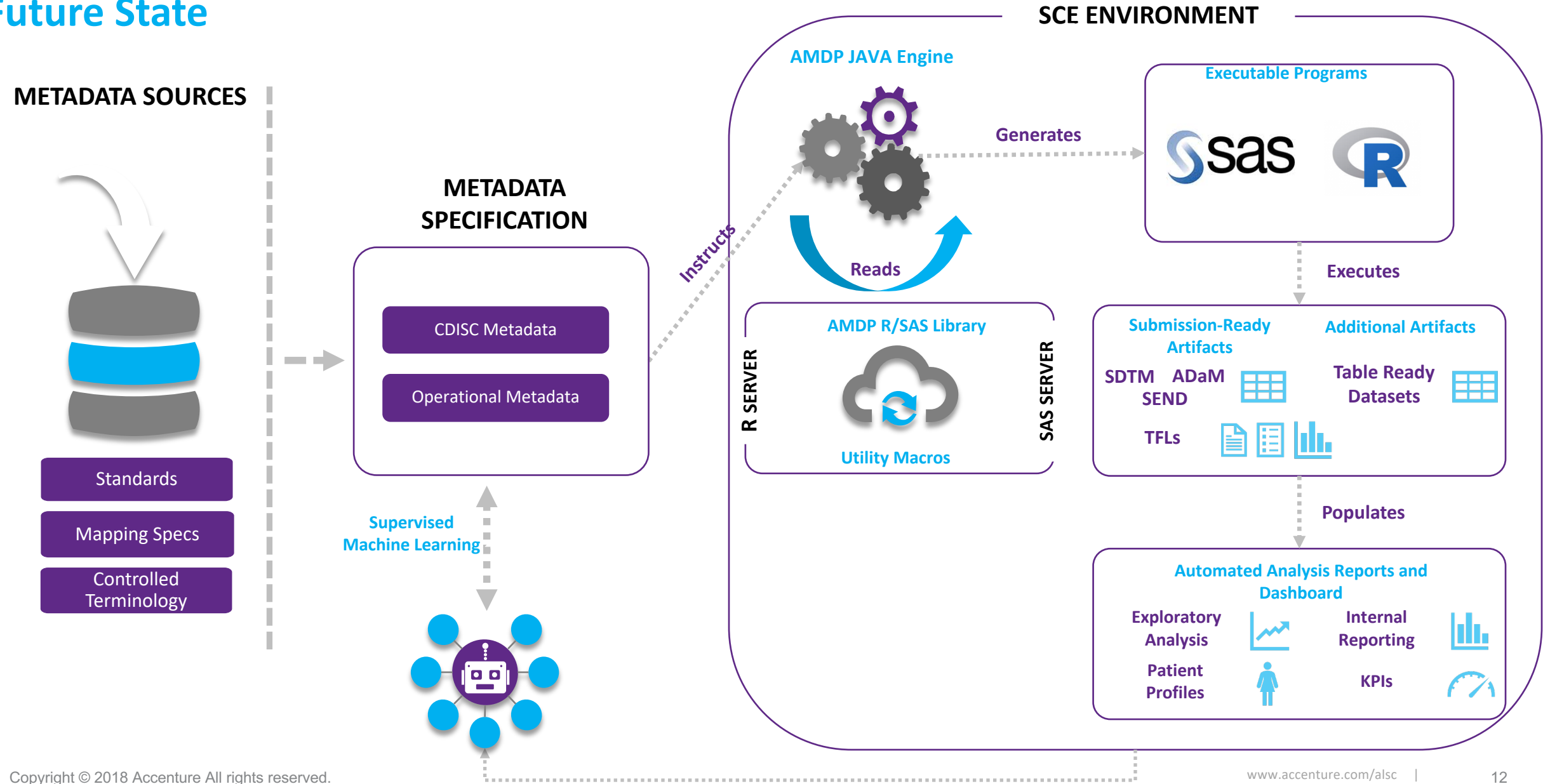
Work Type	Reduction Expected?	Relative Effort (%)
Global Specification Interpretation	Yes	10%
Client-specific Team Training	Yes	5%
Human-readable Specification	Yes	15%
Development		
Machine-readable Specification Development	Yes	40%
AMDP Code Development	Yes	5%
QC and Compliance Reviews	No	15%
Client and Project Management	No	10%
Total		100%

Takeaways

- Time to perform most tasks expected to decrease with study volume
- Very little Code Development
- 55% of effort devoted to specification development

AMPD ARCHITECTURE

Future State



RECAP

- Described an approach and system for the automation of submission-ready artifacts using a metadata-driven code generating engine
- Discussed the validation of the system
- Presented results from a pilot study to convert source data to 48 SDTM domains
- Proposed enhancements to the system

Questions?

