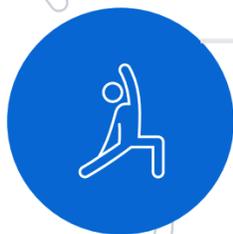


利用自动化工具辅助不同标准的临床试验 数据准备

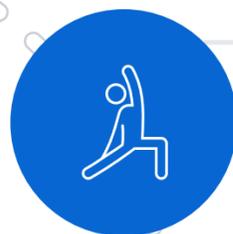
SAS Freddy
2024 Dec





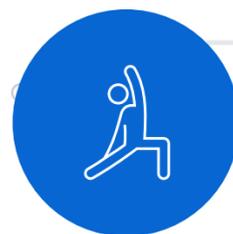
激活元数据以提高数据应用价值

实现主动元数据管理



利用无缝数据流实现数据自动化编排

借助Viya工作站功能



SAS在生命科学端到端服务能力

统一平台提高研发效率

激活元数据以提高数据应用价值

实现主动元数据管理



临床研究元数据标准

元数据三种常见的分类方式



CDISC中的元数据标准及数据/程序管理文件

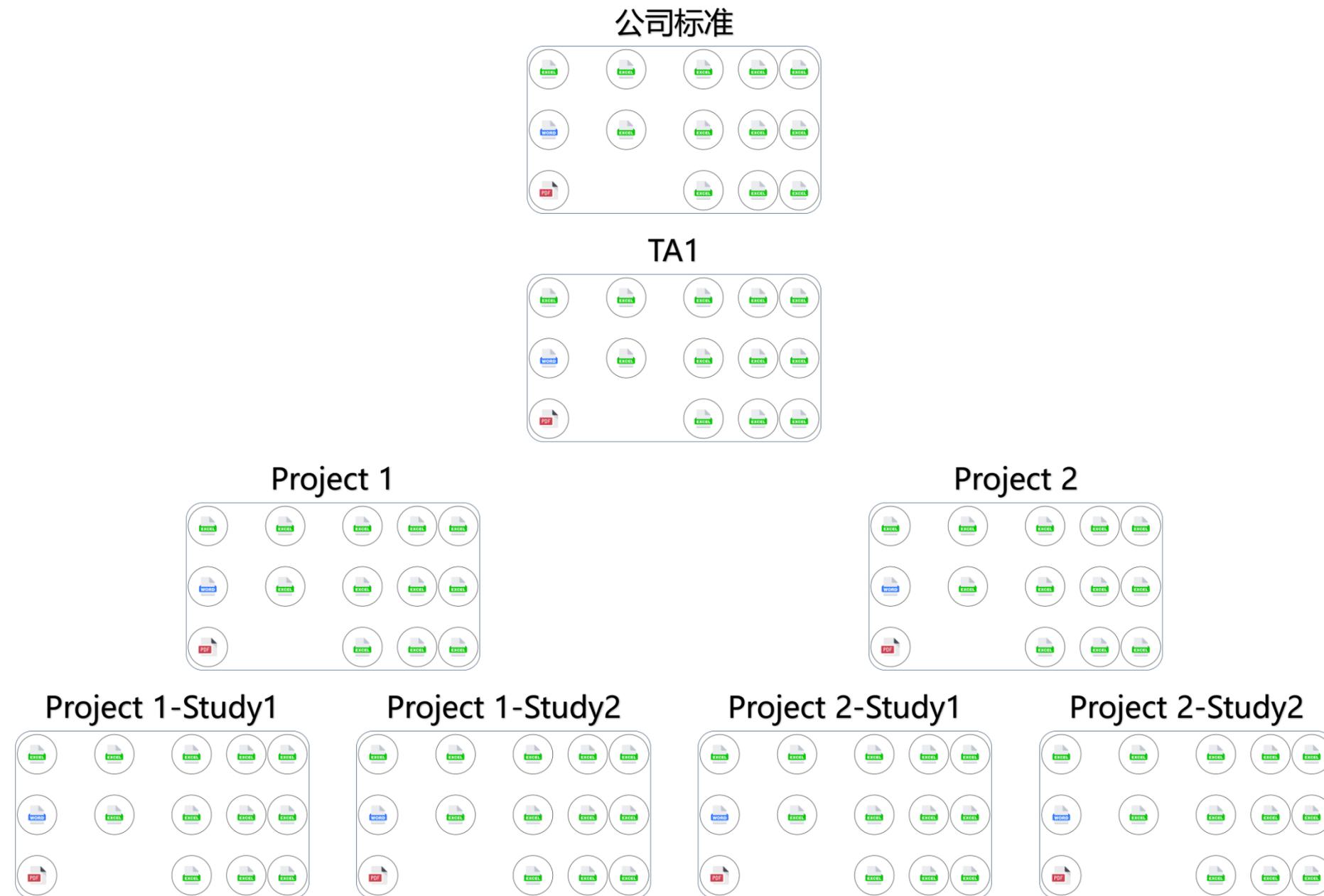
共12个spreadsheets, 1份 Word文件及1份PDF文件



来源: PhUSE 2017 Paper SI02

现阶段相互独立的元数据文件导致工作效率的降低

被动的元数据管理导致数据利用率低



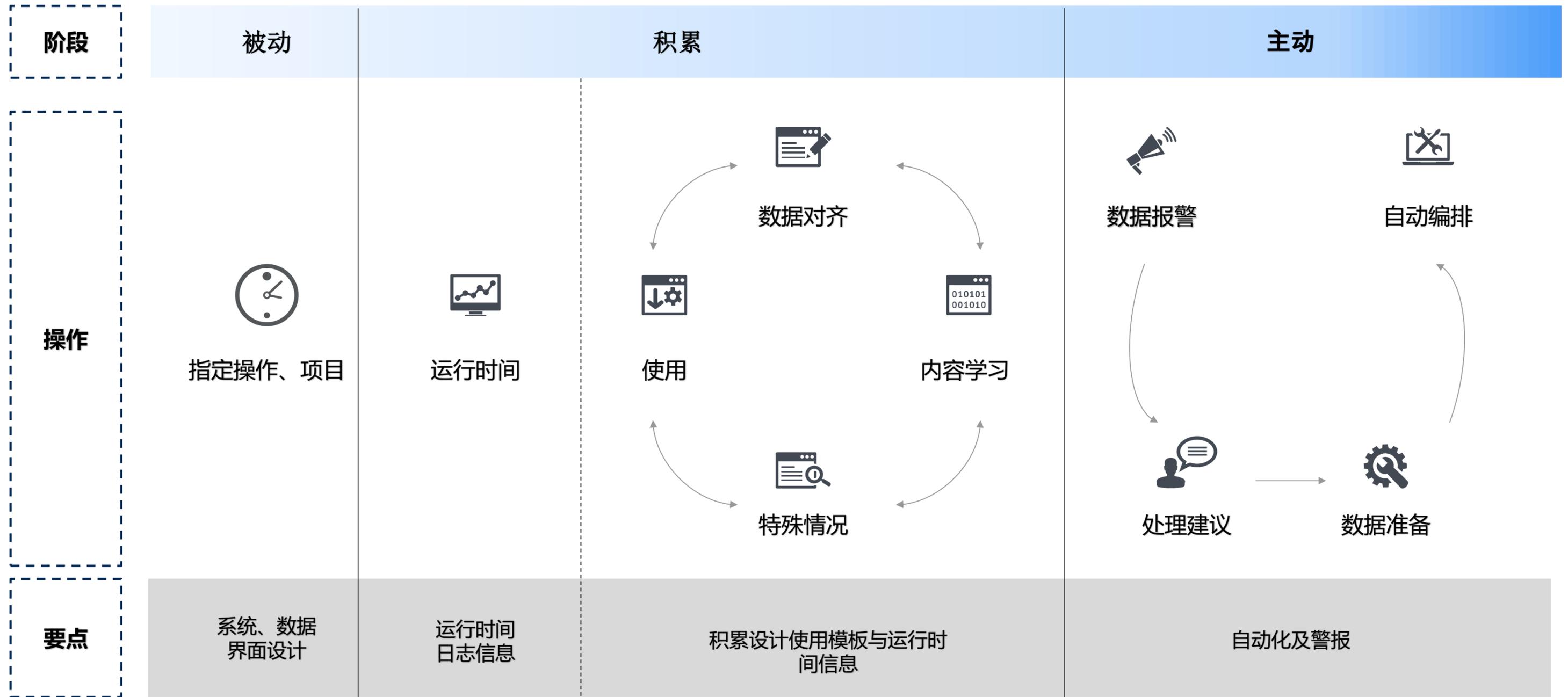
来源: PhUSE 2017 Paper SI02

目前的未满足需求

- 每项研究需要多个电子表格, 导致断开的元数据出现数据孤岛
- 元数据不同元素之间的关系难以理解
- 难以直接从元数据开始进行编程分析
- 难以让多个用户在同一个文件上进行统一协作, 以及很难让不同的用户理解自己部分的修改会如何影响其他人
- 需要使用额外工具来进行CDISC合规和标准管理的评估
- 不透明的元数据管理系统会导致内置数据冗余和缺乏可追溯性

从被动到自动、智能化地主动元数据管理技术

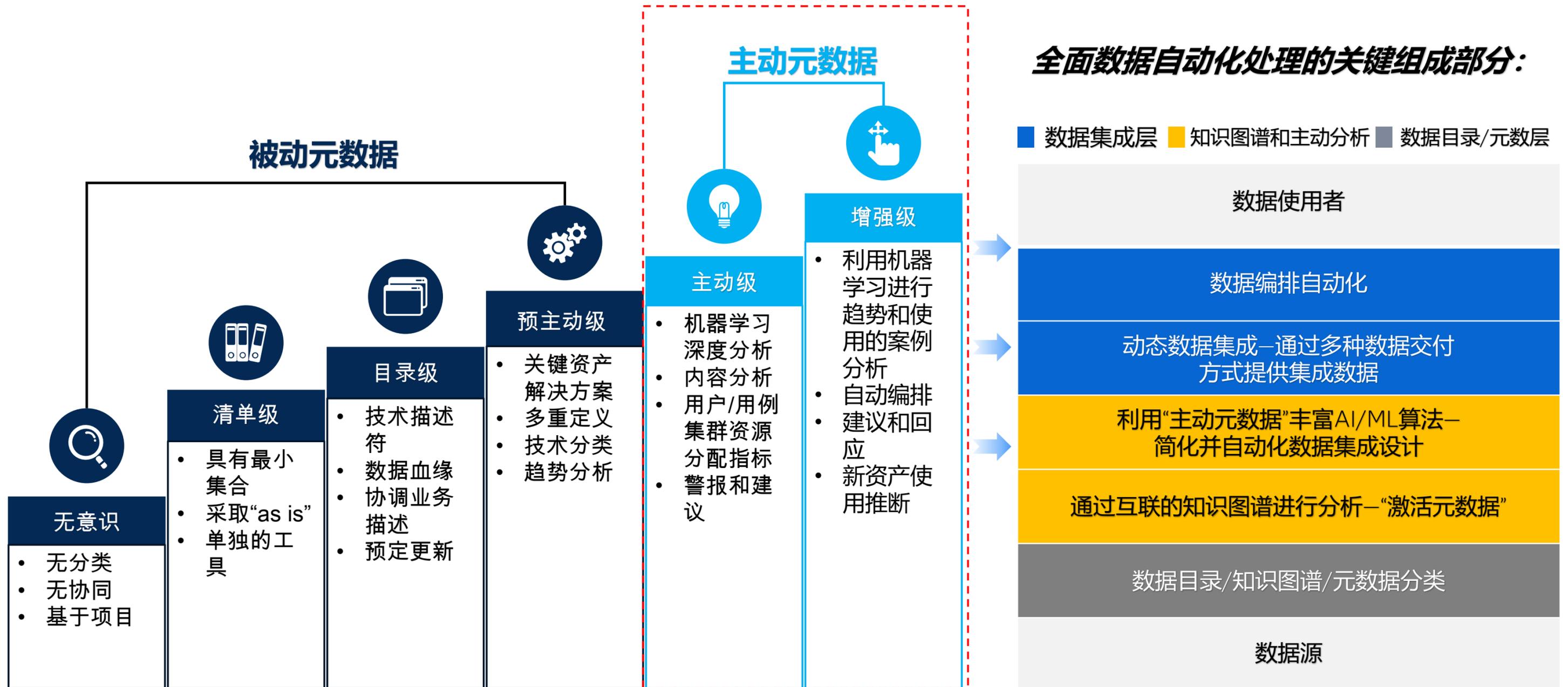
全面、精准、智能



来源: Gartner

实现元数据的主动管理将为用户带来极大便利

主动元数据管理技术逐渐成熟



来源: Gartner

利用无缝数据流实现数据自动化编排

帮助CDISC新标准在企业内实施



SAS Viya, 一个平台, 多个功能界面

可视化界面及编程界面

SAS Viya Platform

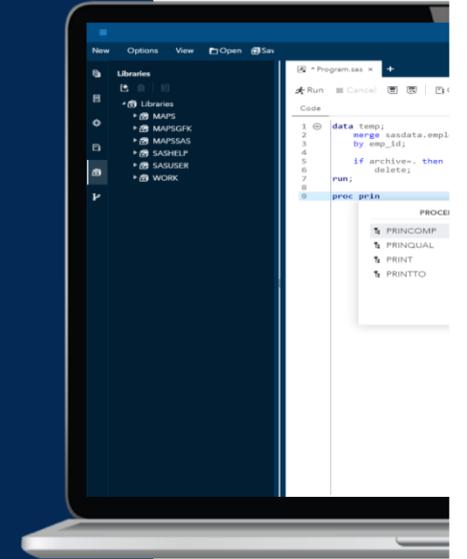
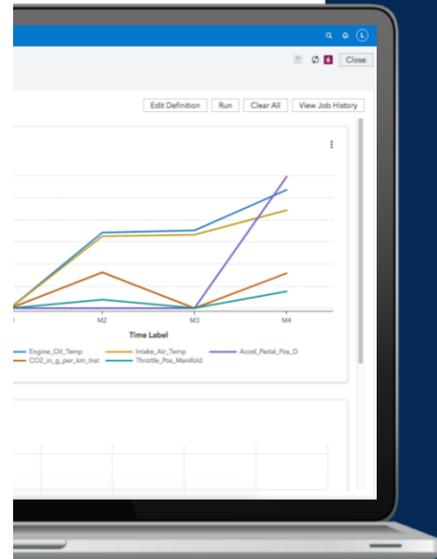
两种主要页面支持多种类型的SAS Viya用户:

可视化界面

更适用于喜欢使用
拖拉拽作为主要工
作模式的用户

编程界面

更适用于SAS编程
基础较好的用户



SAS Information Catalog-资产发现

元数据管理平台

- SAS Information Catalog: 一个直观的分析 and 人工智能自助服务发现门户, 用户可以在开始后续SDTM及ADaM处理之前探索数据资产的质量, 以提高数据可用性

主要功能

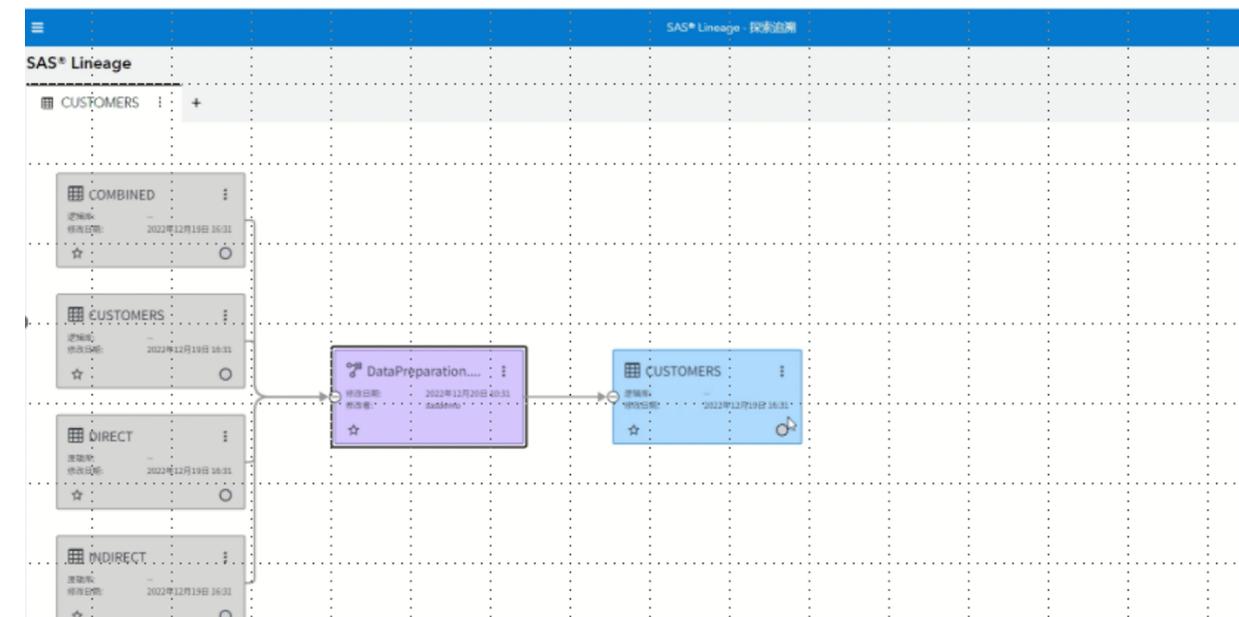
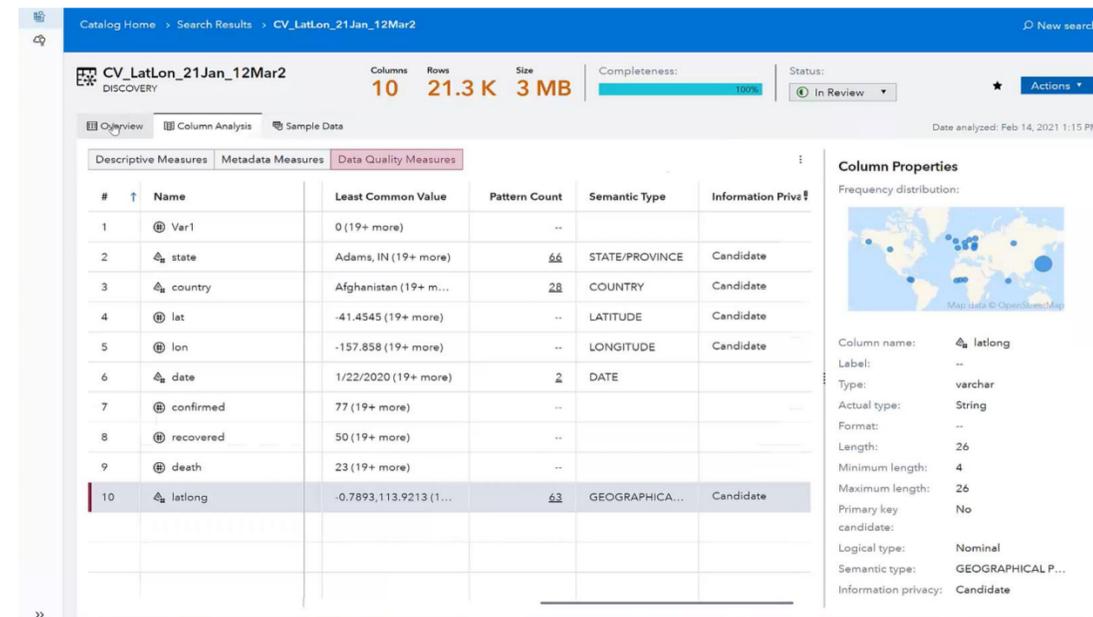
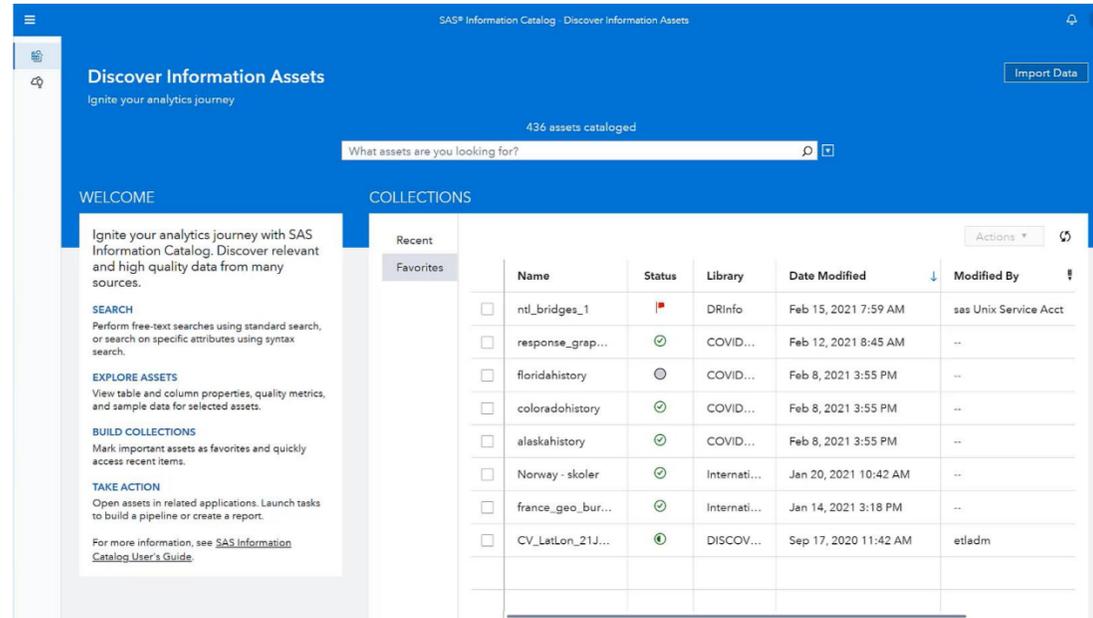
- 分析数据以评估数据质量, 同时确保机密数据的治理和监控
- 自动发现数据和分析资产质量问题
- 使用数据来源追踪和元数据管理进行影响分析

The screenshot displays the SAS Information Catalog interface. At the top, it says "Discover Information Assets" and "Ignite your analytics journey". A search bar is present. Below this, a "CATALOG AT A GLANCE" section shows various asset counts: Total assets (654), Datasets (382), Data plans (251), Studio flows (13), Reports (6), and Lookup tables (2). A "WELCOME" message is also visible. The main section is titled "COLLECTIONS" and features a table with columns: Name, Status, Library, Date Created, Cre..., and Asset Type. The table lists several collections, including 2001MOVIES, BURROWS, CUSTOMER_ORDER_SUMMARY, WATER_CLUSTER, and DQ_CASE_SPLIT_REMOVE. The interface also includes "Recent" and "Favorites" filters on the left and an "Import Data" button on the right.

Name	Status	Library	Date Created	Cre...	Asset Type
2001MOVIES	●	COMM...	Mar 11, 2022 6:17 AM	sas...	In-mem...
BURROWS	●	SASHELP	Oct 11, 2021 9:05 PM	sas ...	SAS table
CUSTOMER_ORDER_SUMMARY	●	CATAL...	May 9, 2018 2:28 PM	--	CAS table
WATER_CLUSTER	●	Samples	Mar 11, 2022 8:26 AM	--	CAS table
DQ_CASE_SPLIT_REMOVE	--		Mar 11, 2022 6:57 AM	sas...	Data plan

更容易发现元数据关系及修改调整历史

资产发现和探索追溯界面展示



Data Management in SAS 9

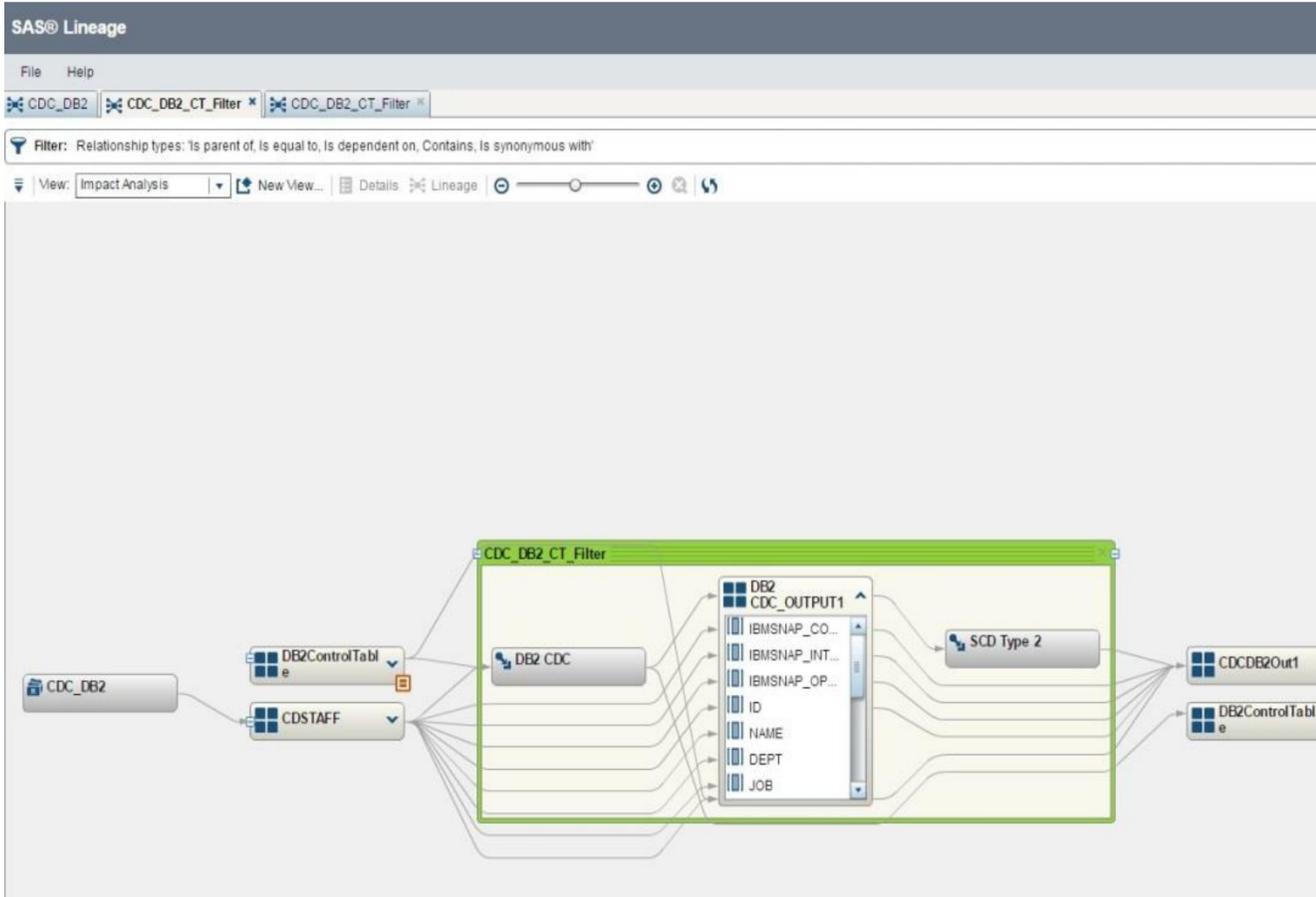
SAS 9中也具有类似的工具功能

The screenshot shows the SAS Business Data Network interface. On the left is a navigation pane with categories like 'All Terms', 'Notifications', and 'My Drafts'. The main area displays a table of terms:

Term Name	Description	Type	Importance	Stat
ActualEducationalSettings	This is a summary of actual educational settings of...	Default	★★★★☆	Not Spec
ActualGradeList	This is a summary of actual grades instructed by a...	Default	★★★★☆	Not Spec
ActualGradeListSearchable	This is a summary of actual grades instructed by a...	Default	★★★★☆	Not Spec
Address_Type	Table that contains address type information from t...	Default	★★★★☆	Not Spec
AddressTypeCode	Address type code.	Default	★★★★☆	Not Spec
AddressTypeDesc	Provides a description of the address type.	Default	★★★★☆	Not Spec
AddressTypeId	Unique identifier for each address type code.	Default	★★★★☆	Not Spec
AddressTypeName	Name of address type.	Default	★★★★☆	Not Spec
AgreementNumber	This is the agreement number (formerly referred to...	Default	★★★★☆	Not Spec
AuthorizedEducationalSettings	This is a summary of authorized educational settin...	Default	★★★★☆	Not Spec
AuthorizedGradeList	This is a summary of the grades authorized to be i...	Default	★★★★☆	Not Spec
AuthorizedGradeListSearchable	This is a summary of the grades authorized to be i...	Default	★★★★☆	Not Spec
AYPDate	Not used.	Default	★★★★☆	Not Spec

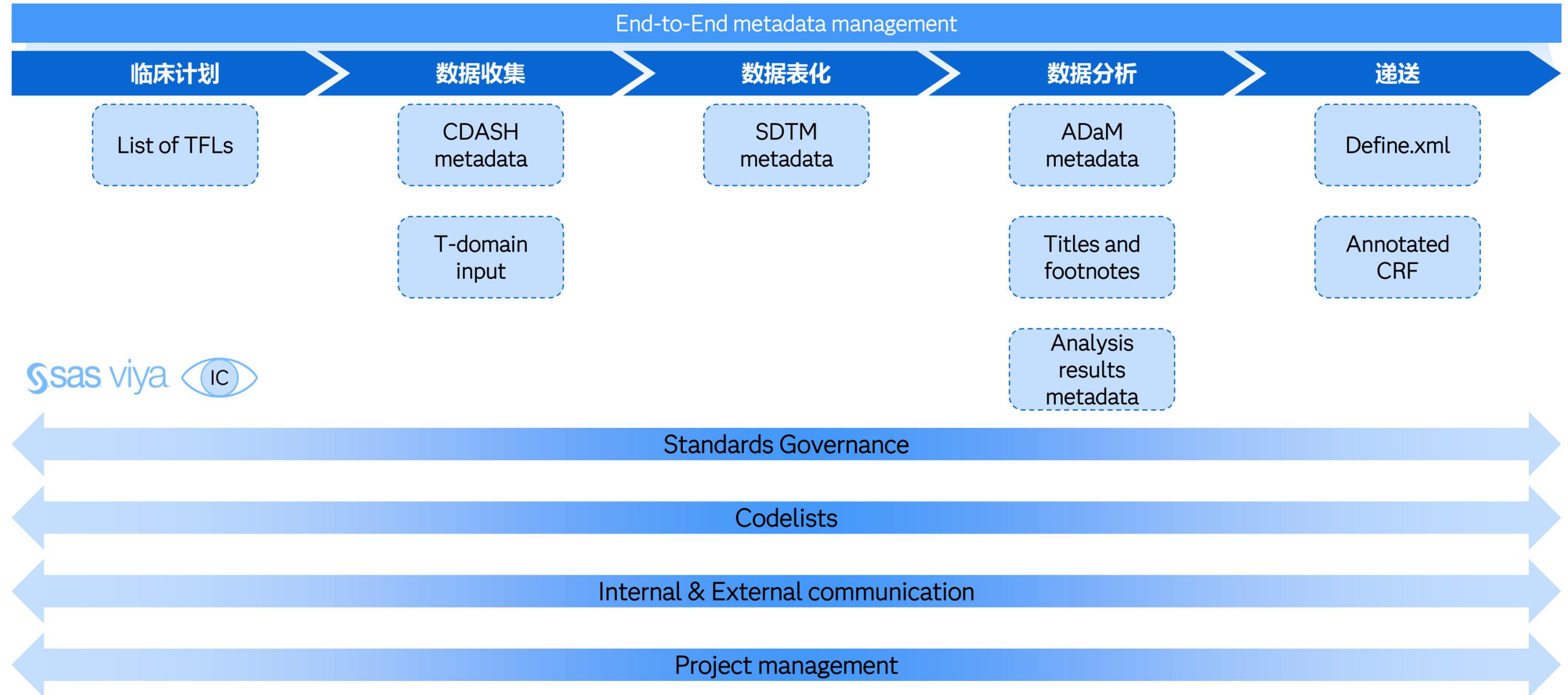
Below the table, the 'ActualEducationalSettings' term is selected, showing its 'Description' in a detailed view:

Description
 This is a summary of actual educational settings offered by an individual entity or a group of entities within a parent entity. This characteristic is derived by the system from the checked actual educational settings. This characteristic can have multiple values. Selectable values: General Education (GenEd); Grades K-12. A public elementary/secondary school that does NOT focus primarily on vocational special or alternative education although it may provide these programs in addition to a regular curriculum. -- Pre-Kindergarten (PK): A national technical term for a state local or privately funded preschool program for children one or two years before kindergarten entry. Pre-K associated with ISD/LEA districts must be reported in EDEN. -- Alternative Education (Alt): An entity that addresses the needs of students that typically cannot be met in a regular school program. The entity provides nontraditional education; serves as an adjunct to a regular school; and falls outside the categories of regular



将不同阶段的metadata进行统一管理

利用资产发现和探索追溯进行更好的元数据管理



来源: PhUSE 2017 Paper S102

FDA及PMDA 要求的数据标准

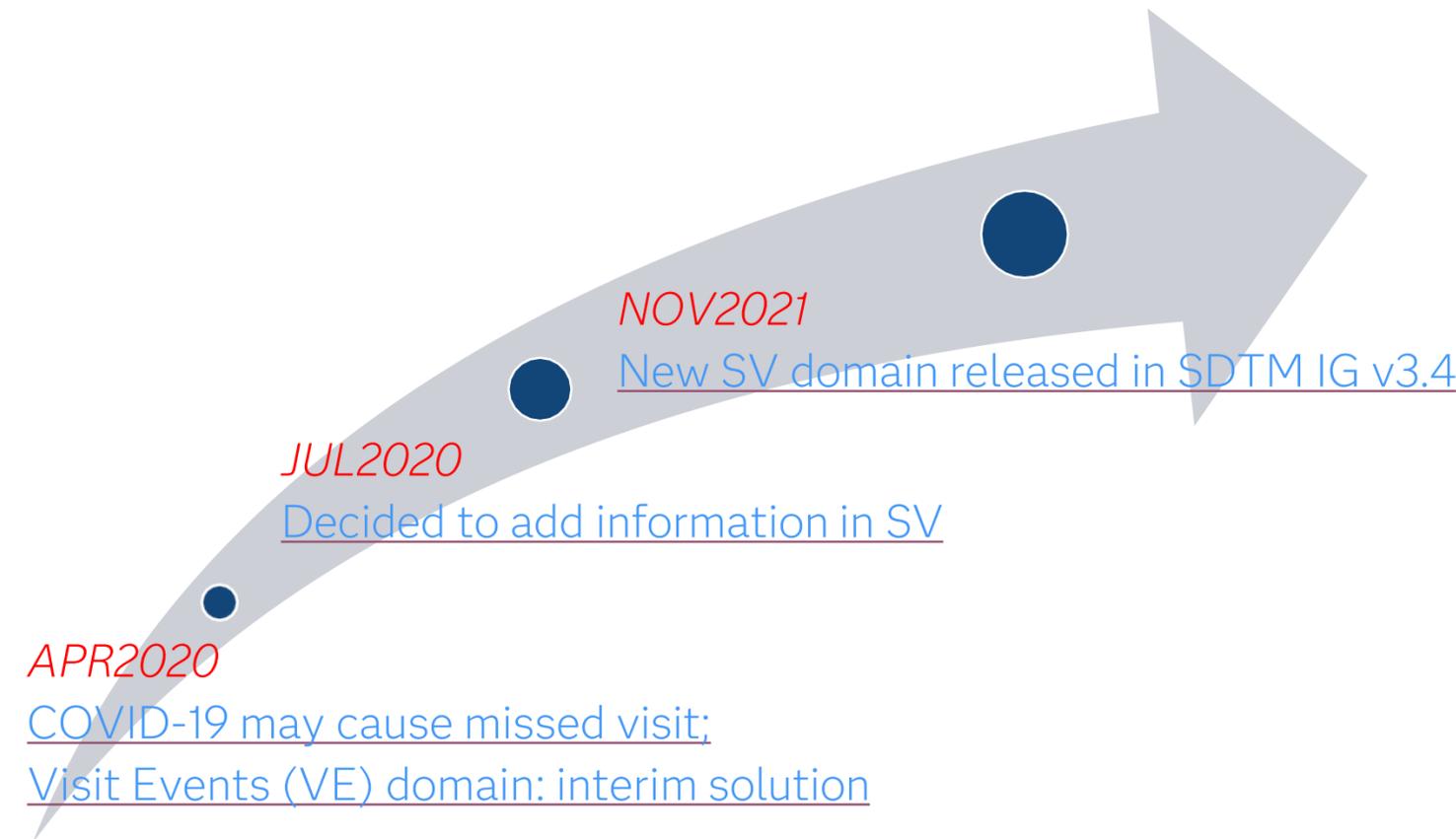
FDA将于2025年3月正式使用SDTMIG 3.4

HA	Version	Date (mandatory)	Source
FDA	SDTMIG 3.3	15Mar2023	FDA Data Stds Catalog_8.01.2022 (v8.2)
PMDA	SDTMIG 3.3	01Apr2023	PMDA Data Stds Catalog (28.02.2023)
FDA	SDTMIG 3.4	15Mar2025	FDA Data Stds Catalog_04.12.2024 (v10.3)

- For FDA Start date is assumed as **First Patient First Visit** on or after the specified date.
- For PMDA Start date is assumed as **submission date**, Study submission on or after the specified date can be done on SDTM IG v3.3 and SDTM IG v3.2 as well until end date is provided.

直接导入新增的Domain及变量的元数据名称及变化规则

以Subject Visits (SV)的变化为例



SV – Specification

sv.xpt, Subject Visits — Special Purpose. One record per actual or planned visit per subject, Tabulation.

Variable Name	Variable Label	Type	Controlled Terms, Codelist or Format ¹	Role	CDISC Notes	Core
STUDYID	Study Identifier	Char		Identifier	Unique identifier for a study.	Req
DOMAIN	Domain Abbreviation	Char	SV	Identifier	Two-character abbreviation for the domain most relevant to the observation. The domain abbreviation is also used as a prefix for variables to ensure uniqueness when datasets are merged.	Req
USUBJID	Unique Subject Identifier	Char		Identifier	Identifier used to uniquely identify a subject across all studies for all applications or submissions involving the product.	Req
VISITNUM	Visit Number	Num		Topic	Clinical encounter number. Numeric version of VISIT, used for sorting.	Req
VISIT	Visit Name	Char		Synonym Qualifier	Protocol-defined description of a clinical encounter.	Perm
SVPRESP	Pre-specified	Char	(NY)	Variable Qualifier	Used to indicate whether the visit was planned (i.e., visits specified in the TV domain). Value is "Y" for planned visits, null for unplanned visits.	Exp
SVOCCUR	Occurrence	Char	(NY)	Record Qualifier	Used to record whether a planned visit occurred. The value is null for unplanned visits.	Exp
SVREASOC	Reason for Occur Value	Char		Record Qualifier	The reason for the value in SVOCCUR. If SVOCCUR="N", SVREASOC is the reason the visit did not occur.	Perm
SVCNTMOD	Contact Mode	Char	(CNTMODE)	Record Qualifier	The way in which the visit was conducted. Examples: "IN PERSON", "TELEPHONE CALL", "IVRS".	Perm
SVEPCHGI	Epi/Pandemic Related Change Indicator	Char	(NY)	Record Qualifier	Indicates whether the visit was changed due to an epidemic or pandemic.	Perm
VISITDY	Planned Study Day of Visit	Num		Timing	Planned study day of VISIT. Should be an integer.	Perm
SVSTDTC	Start Date/Time of Observation	Char	ISO 8601 datetime or interval	Timing	Start date/time of an observation represented in ISO 8601 character format.	Exp
SVENDTC	End Date/Time of Observation	Char	ISO 8601 datetime or interval	Timing	End date/time of the observation represented in ISO 8601 character format.	Exp
SVSTDY	Study Day of Start of Observation	Num		Timing	Actual study day of start of observation expressed in integer days relative to the sponsor-defined RFSTDTC in Demographics.	Perm
SVENDY	Study Day of End of Observation	Num		Timing	Actual study day of end of observation expressed in integer days relative to the sponsor-defined RFSTDTC in Demographics.	Perm
SVUPDES	Description of Unplanned Visit	Char		Record Qualifier	Description of what happened to the subject during an unplanned visit. Only populated for unplanned visits.	Perm

SV修改了整个域的结构，新增了5个变量。结构：包括受试者实际的和计划的访视。

在SDTMIG 3.3中仅需要实际的访视。因为SV结构调整，同时也加入了三个相关的变量：

- SVPRESP：表示是否方式是计划的。
- SVOCCUR：表示计划的访视是否发生。
- SVREASOC：表示是否发生 (SVOCCUR) 的原因。如果未发生，则表示的是未发生的原因。

同时受到新冠疫情大流行的影响，对临床试验的运行产生了较大的冲击。在SV中加入了两个新变量，用于区分：

- SVCNTMOD：访问的方式，比如“当面”，“电话”
- SVEPCHGI：访视是否因为大流行而变化

aCRF到SDTM映射工作的重要性及目前面临的挑战

较为依赖人工且失误难以避免

➤ 重要性

- aCRF到SDTM映射是提交临床试验数据集必需的工作
- 映射工作直接影响了数据分析和结果报告的正确性
- 映射正确与否关系到最后研究结果的合规性和可靠性

➤ 面临的挑战

- 映射工作目前主要依靠人工，非常费时费力
- 不同研究的CRF格式各异，映射难度大且重复性工作多
- 映射工作需要人工持续回顾核验，容易出现人为失误
- 随着SDTM IG版本的迭代，部分同事在没有完全适应3.3的情况下需要同时处理一些需要使用3.4的项目

SAS Viya Studio统一集成开发环境

集代码开发， 过程流设计于一体

SAS Enterprise Guide



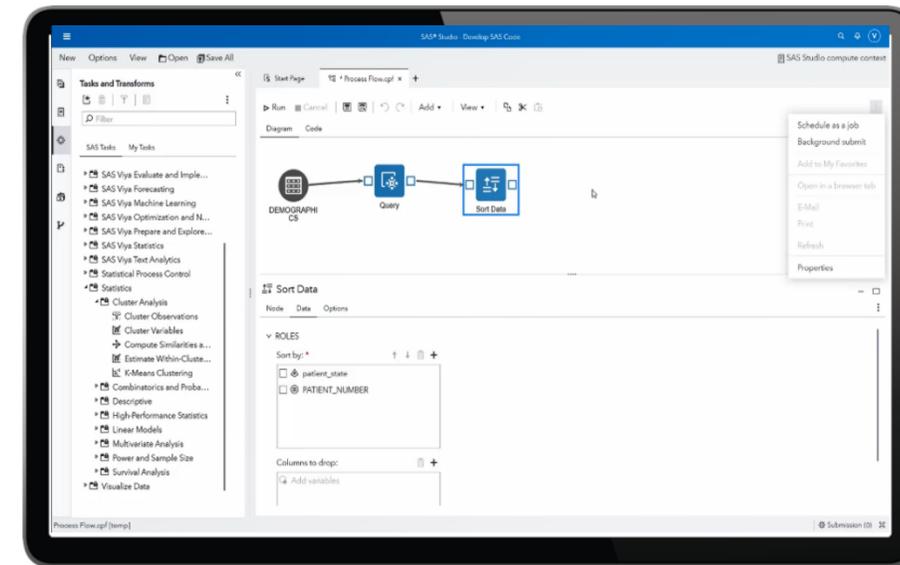
SAS Data Integration Studio



SAS Code, Servers & Libraries



SAS Studio



Code

功能强大的 SAS 代码编辑器，具有自动完成功能和语法帮助功能

Design

任务、查询和流

Collaborate

GIT集成、共享任务和代码段

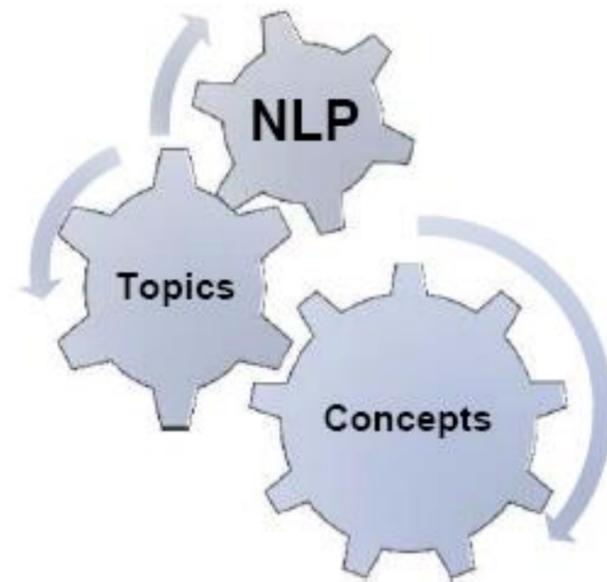
SAS Virtual Text Analysis 可视化文本分析

将文本挖掘、上下文提取、分类、情感分析和智能搜索结合在灵活的框架中

电子病历案例 (非结构化数据)

Doctor Status Note:
Deg juj to koiuklk pert kophgsmos opoper aepwer mim, sjsæjdkesk
Juuk moweri jej kek mågdfufd **fever 38.9** and **pain when urinating**.
Figh babe 20 jyj, kelolike ukiukjuu **patient preveiusly had catheter** nandneftge.

Trigger-algorithm "fever"



Trigger-algorithm "UTI pain"



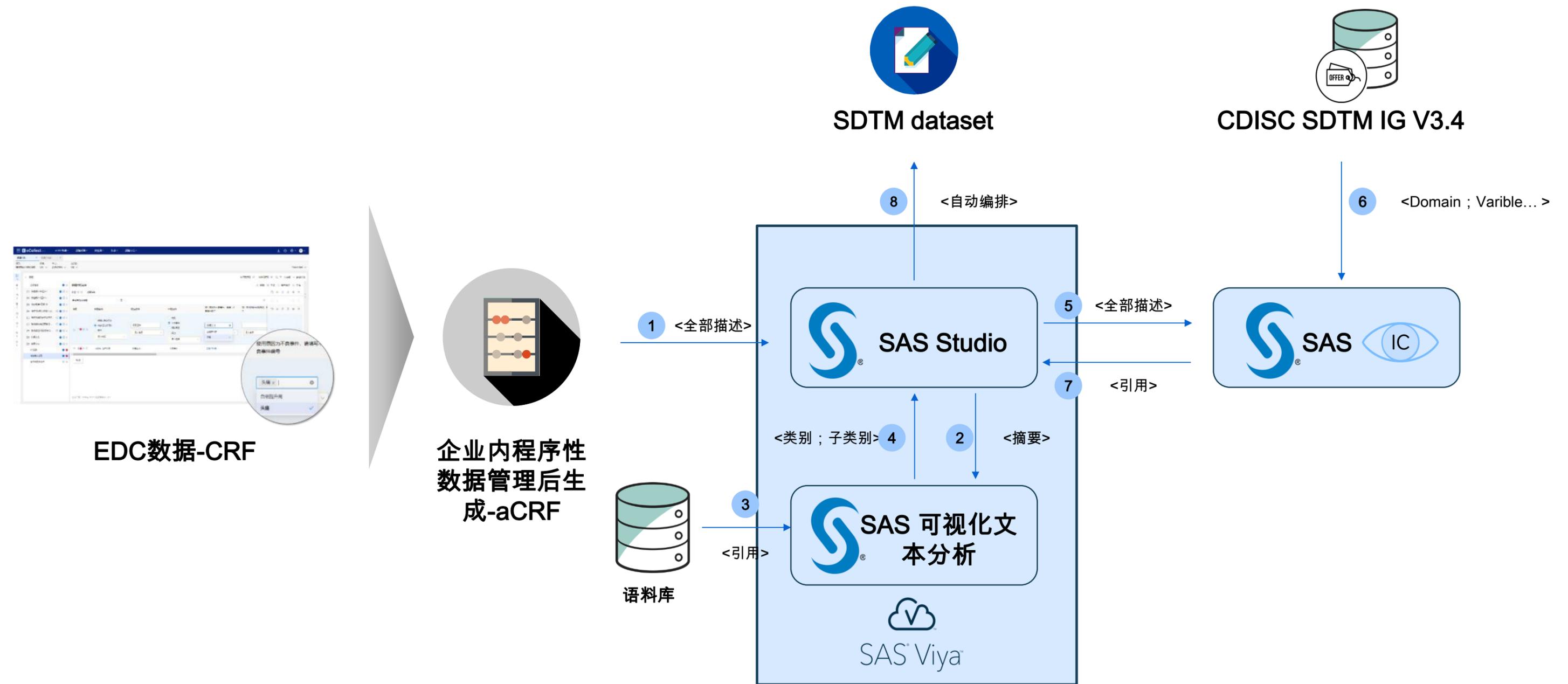
转换非结构化文本
通过文本分析算法 (触发Liti算法, 布尔运算等)
形成结构化数据

PID (unique person ID)	Date	Trigger code * (= fever)
PID (unique person ID)	Date	Trigger code *(= uti pain)

生成结构化数据

借助SAS Studio完成数据标准化的流程

借助SAS过程流及可视化文本分析



利用这一过程流进行自动化编排

根据aCRF模板自动调整过程流

DM = Demographics

DEMOGRAPHICS (dem)

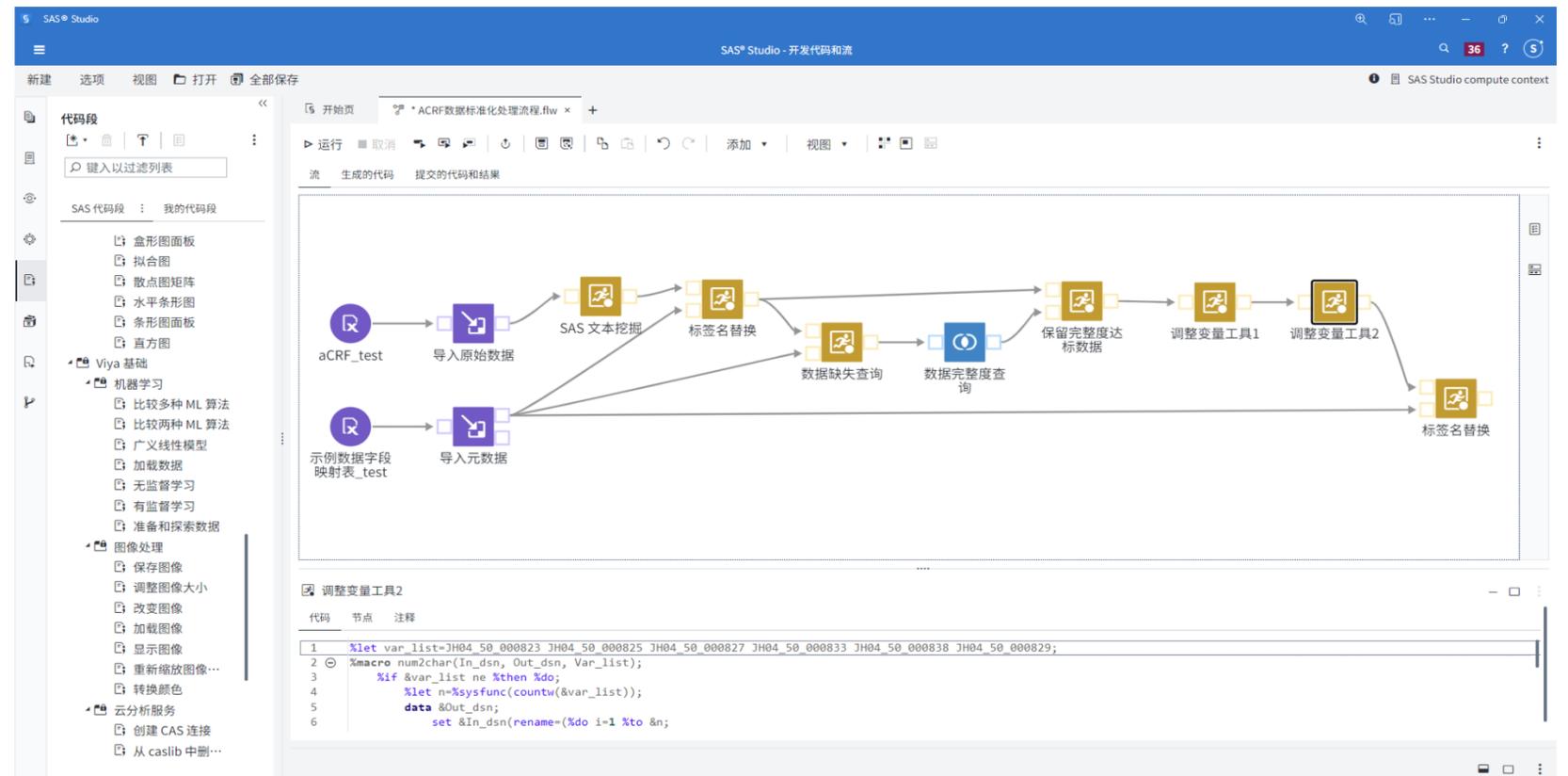
1. Birth date (read-only) [Birth date]	[Birth date] / Req/Req / Req (1910-2010)	BRTHDTC
2. Sex (read-only) [Sex]	[Sex] <input type="radio"/> Male <input type="radio"/> Female	SEX
3. Ethnicity 1 [Ethnicity 1]	[Ethnicity 1] <input type="radio"/> Hispanic or Latino <input type="radio"/> Not Hispanic or Latino	ETHNIC
4. Ethnicity 2 [Ethnicity 2]	[Ethnicity 2] <input type="radio"/> Japanese <input type="radio"/> Not Japanese	ETHNIC2 in SUPDM
5. Race [Race]	[Race] [White] [Black or African American] [Black or African American] [Asian] [American Indian or Alaska Native] [American Indian or Alaska Native] [Native Hawaiian or Other Pacific Islander] [Native Hawaiian or Other Pacific Islander] [Not collected at this site] [Other, specify]: A200 =	RACE RACENC in SUPDM where RACE = "" RACEOTH in SUPDM where RACE = "OTHER"

Key: [*] = Source verification required [*] = AS22 Only
Note: Source verification critical settings made in Inform will override any settings made in Control Designer.

- aCRF 表

PORTAL_DATASET	PORTAL_VARIABLE	PVAR_CODELIST_NAME	PVAR_ORIGIN	SRCMORIGIN	ALIAS	PVAR_INTENDED_DATA_TYPE	PVAR_INTENDED_LENGTH	PVAR_INTENDED_LABEL	SEQ_NO	SUPP_FLAG	CORE	COMM_FLAG
1	RS	STUDYID	Protocol	Assigned		Character	200	Study Identifier	1	Req		
2	RS	DOMAIN	Domain	Assigned		Character	2	Domain Abbreviation	2	Req		
3	RS	USUBJID	Unique Subject Identifier	Derived		Character	60	Unique Subject Identifier	3	Req		
4	RS	RSSEQ	Sequence Number	Derived		Number	8	Sequence Number	4	Req		
5	RS	RSGRPID	Group ID	eDT.Assigned		Character	100	Group ID	5	Perm		
6	RS	RSREFID	Reference ID	eDT		Character	200	Reference ID	6	Perm		
7	RS	RSSPID	Sponsor-Defined Identifier	eDT.CRF		Character	200	Sponsor-Defined Identifier	7	Perm		
8	RS	RSLNKID	Link ID	Assigned		Character	200	Link ID	8	Perm		
9	RS	RSLNKGRP	Link Group	Assigned		Character	200	Link Group	9	Perm		
10	RS	RSTESTCD	Response Assessment Short Name	Derived.Assigned		Character	8	Response Assessment Short Name	10	Req		
11	RS	RSTEST	Response Assessment Name	Assigned.eDT		Character	40	Response Assessment Name	11	Req		
12	RS	RSCAT	Category for Response Assessment	Assigned		Character	80	Category for Response Assessment	12	Exp		
13	RS	RSSCAT	Subcategory	Assigned		Character	200	Subcategory	13	Perm		
14	RS	RSORRES	Response Assessment Original Result	CRF.Assigned.eDT		Character	200	Response Assessment Original Result	14	Exp		
15	RS	RSORRESU	Original Units	Assigned.CRF.eDT		Character	200	Original Units	15	Perm		
16	RS	RSSTRES	Response Assessment Result in Std Format	Assigned		Character	200	Response Assessment Result in Std Format	16	Exp		
17	RS	RSSTRESN	Numeric Result/Finding in Standard Units	Assigned		Number	8	Numeric Result/Finding in Standard Units	17	Perm		
18	RS	RSSTRESU	Standard Units	Assigned		Character	200	Standard Units	18	Perm		
19	RS	RSRESCAT	Result Category	CRF.eDT		Character	40	Result Category	19	Perm		
20	RS	RSSTAT	Completion Status	Assigned		Character	8	Completion Status	20	Perm		
21	RS	RSREASND	Reason Response Assessment Not Performe	CRF.Derived.eDT		Character	200	Reason Response Assessment Not Performe	21	Perm		
22	RS	RSNAM	Vendor Name	Assigned		Character	200	Vendor Name	22	Perm		
23	RS	RSLOC	Location Used for the Measurement	CRF.Assigned.eDT		Character	200	Location Used for the Measurement	23	Perm		
24	RS	RSLAT	Laterality	eDT.CRF		Character	200	Laterality	24	Perm		
25	RS	RSOR	Directionality	eDT.CRF		Character	40	Directionality	25	Perm		
26	RS	RSMETHOD	Method of Test or Examination	Assigned.CRF		Character	200	Method of Test or Examination	26	Perm		
27	RS	RSBFL	Baseline Flag	Derived		Character	1	Baseline Flag	27	Perm		
28	RS	RSEVAL	Evaluator	Assigned.eDT.CRF		Character	200	Evaluator	28	Exp		
29	RS	RSEVALID	Evaluator Identifier	eDT		Character	40	Evaluator Identifier	29	Perm		
30	RS	RSACPTFL	Accepted Record Flag	Assigned		Character	1	Accepted Record Flag	30	Perm		
31	RS	RSVITNUM	Visit Number	Assigned		Number	8	Visit Number	31	Exp		
32	RS	RSVIT	Visit Name	Assigned		Character	80	Visit Name	32	Perm		
33	RS	RSEPOCH	Epoch	Derived		Character	50	Epoch	33	Perm		

- 字段映射样例表 (metadata)



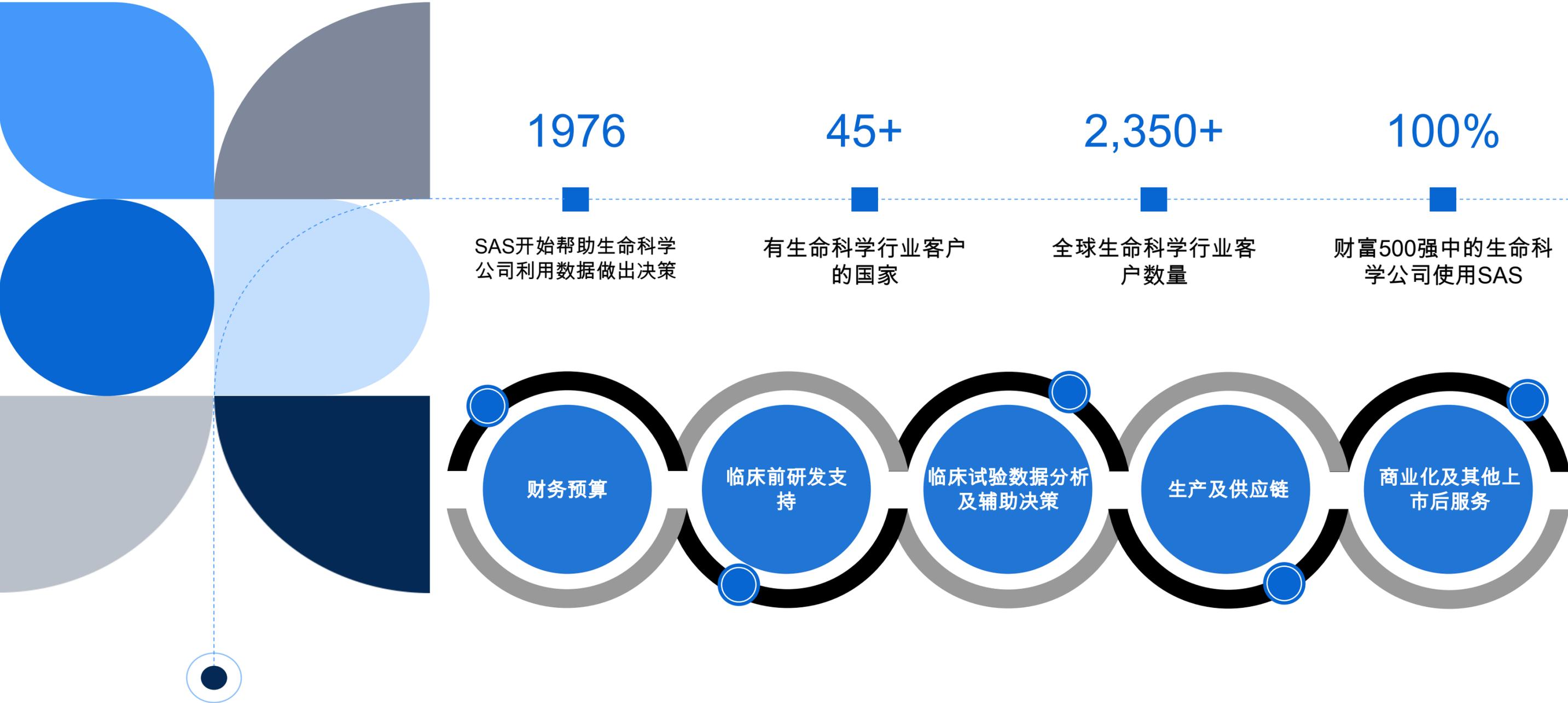
SAS在生命科学端到端 服务能力

统一平台提高研发效率



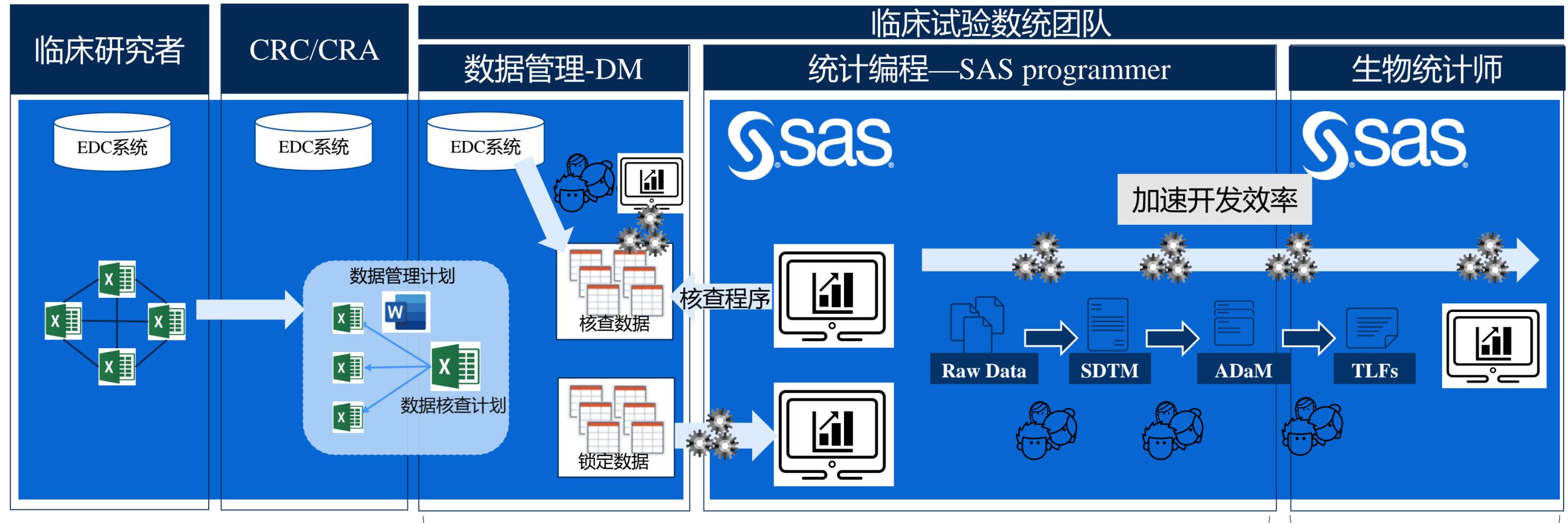
自1976年成立以来，SAS在生命科学领域有着极大影响力

2350+客户数，45+国家



利用一体化化工具加速临床研究效率

解决实际研发中的未满足需求



- DM团队数据核查和质控负担重
- SAS程序员为DM团队编写数据核查脚本任务繁重，调用机制复杂
- CDISC数据管理和元数据管理缺乏系统化标准操作流程
- 缺乏SAS宏程序知识产权管理

- SAP计划编写和分发
- 统计结果实时查看
- 三期适应性模型开发

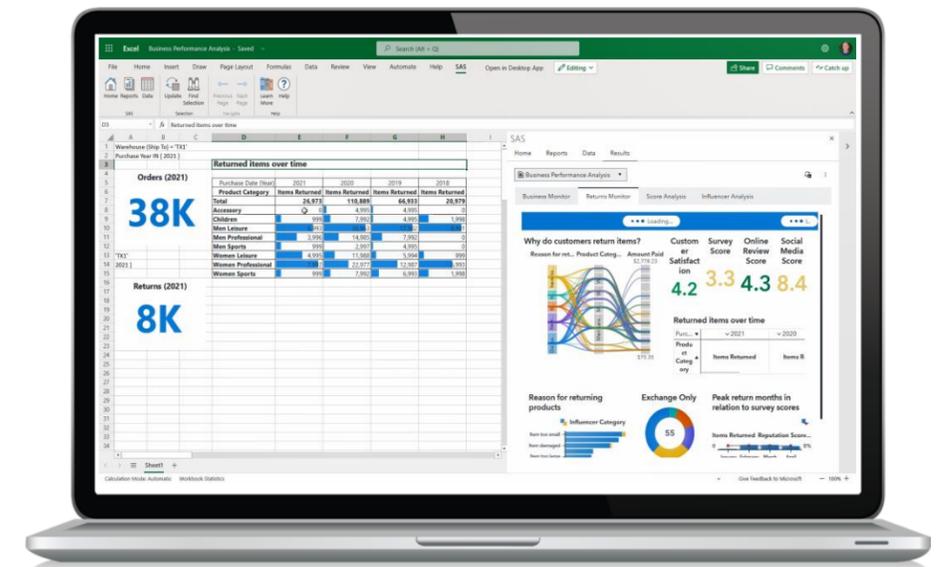
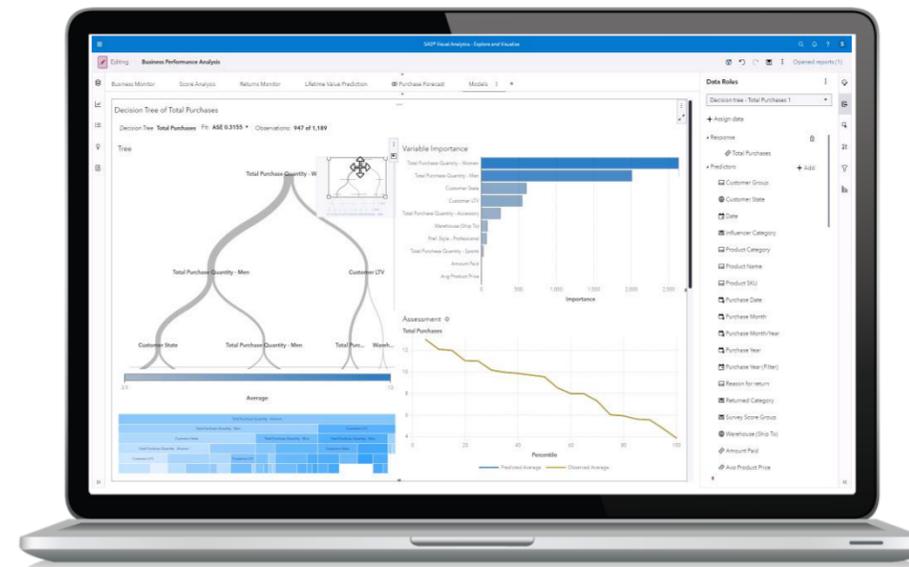
临床试验数据管理与核查分析的可视化技术

提高用户对数据含义的认知

- 列出临床试验数据常用Domain: DM/IC/IE/MH/LB/PE/VS/AE/CM/DS
- 逻辑核查类型
 - 窗口和范围核查
 - 交叉逻辑关系核查
 - 缺失和不完整行数据核查
 - 外源数据核查
 - 其他核查方法

- 基于风险的质量管理
- 医学监察可视化审阅
- 药物警戒信号可视化管理
- 研究数据库实时可视化
- 数据管理可视化
- 真实世界数据交互可视化
- 统计结果TFL可视化

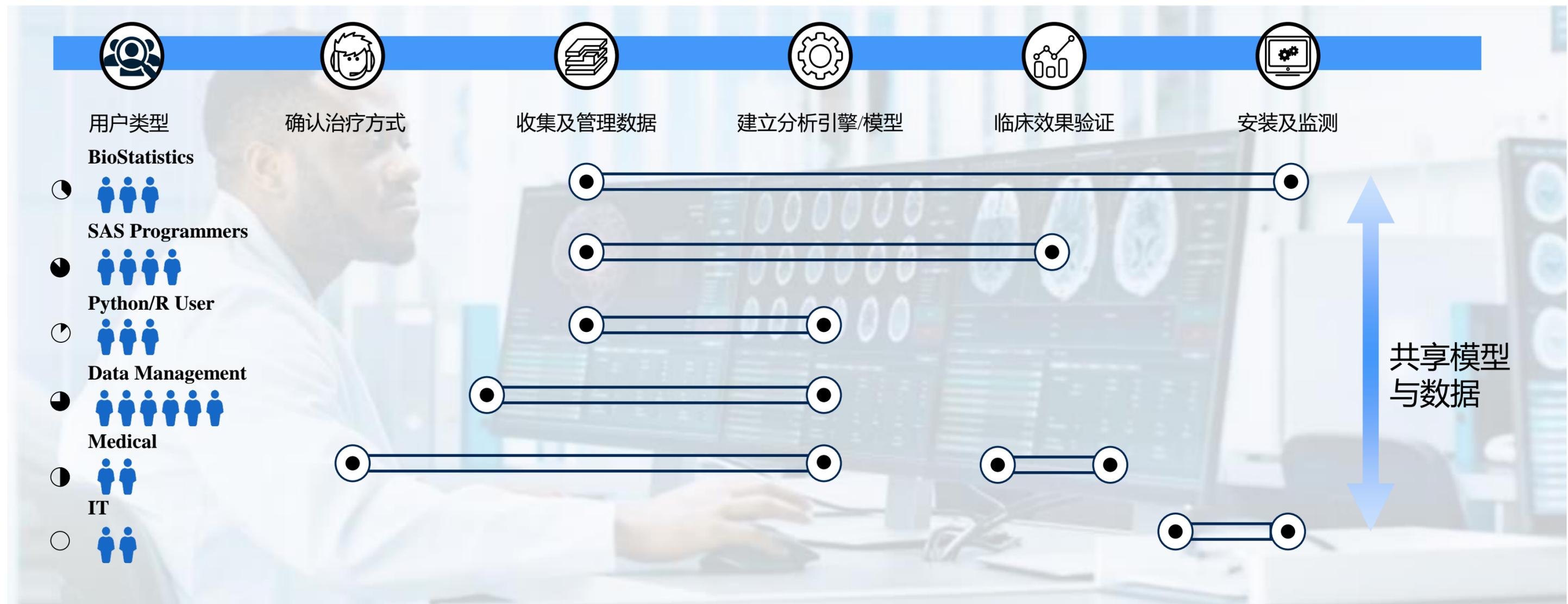
可视化分析技术是充分利用人们的视觉感知能力，增加数据呈现方式以提高人们对数据含义的认知



未来端到端数据科学技术平台的展望

涵盖整个临床开发数据流程，提供一站式解决服务

- 临床试验数据互联互通：数据流转畅通无阻，数据标准业界趋同，数据访问安全可控
- 生物统计应用无处不在：分析工具因人而异，分析场景业务驱动，分析应用灵活轻量
- 人工智能自动化辅助：无缝衔接试验设计与试验实施，自动监控不合规操作，人工智能模型可审计可解释



Thanks

sas.com

