

收录于话题

#数据治理体系 #数据标准

从2024年**全国数据工作会议**上获悉：

一、关于数据标准化

国家数据局局长刘烈宏表示，将建立健全国家数据标准化体制机制，研究成立全国数据标准化技术委员会，统筹指导我国数据标准化工作，加快研究制定一批数据领域国家和行业标准。

刘烈宏说，数据工作是开创性工作，将切实提升创新力，加快数据学科体系和人才队伍建设，选择有较好基础的高校，开展数字经济产教融合合作试点；加强标准引领，印发国家数据标准体系建设指南，指导发布数据要素流通标准化白皮书。针对数据工作面临的众多新情况新问题，刘烈宏说将切实提升研究力，加强理论研究，做好规划编制，抓紧深入谋划“十五五”时期数据工作主要任务，加快组建国家数据咨询专家委员会，发挥智库作用。

二、什么是数据标准？

数据标准是指企业为保障数据的内外部使用和交换的一致性和准确性而制定的规范性约束。

（一）数据标准的内涵

数据标准（Data Standards）是指保障数据的内外部使用和交换的一致性和准确性的规范性约束¹。在数字化过程中，数据是业务活动在信息系统中的真实反映。由于业务对象在信息系统中以数据的形式存在，数据标准相关管理活动均需以业务为基础，并以标准的形式规范业务对象在各信息系统中的统一定义和应用，以提升企业在业务协同、监管合规、数据共享开放、数据分析应用等各方面的能力。

👤 公众号 · 数据治理体系

而数据标准管理则是一套由管理制度、管控流程、技术工具共同组成的体系，是通过这套体系的推广，应用统一的数据定义、数据分类、记录格式和转换、编码等实现数据的标准化。

数据标准管理的目标是通过统一的数据标准制定和发布，结合制度约束、系统控制等手段，实现数据的完整性、有效性、一致性、规范性、开放性和共享性管理，为数据资产管理提供管理依据。

数据标准管理的对象可以分为**数据模型**、**主数据和参考数据**、**指标数据**三大类，每一类均可采用以数据元为数据标准制定的基本单元构建数据标准体系。

1、模型数据标准

基础数据指业务流程中直接产生的，未经过加工和处理的基础业务信息，**模型数据**是指对基础类 数据特征的抽象和描述。

模型数据标准是为了统一企业业务活动相关数据的一致性和准确性，解决业务间数据一致性和数据整合，按照数据标准管理过程制定的数据标准，模型数据标准也是元数据管理的主要内容之一。

以下是某银行为确保数据标准使用，形成的一整套**模型数据标准**的信息项属性架构：

基础属性	信息项编号	标准主题	信息项大类	信息项小类	信息项细类
业务属性	信息项中文名称	信息项英文名称	业务定义	制定依据	业务规则
技术属性	数据类型	数据长度	数据精度	取值范围	度量单位
管理属性	归口管理部门	版本号	制定日期	备注	
代码扩展定义	代码值	代码描述	代码编码规则	公众号 · 数据治理体系	

下图是某运营商数据仓库DWD模型层常用数据元的标准定义示例：

下图是某运营商数据仓库DWD层数据元后缀规范示例：

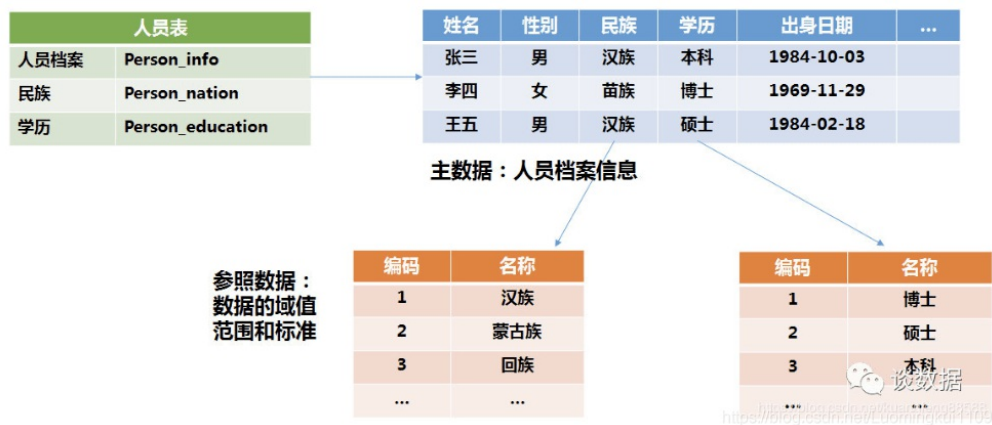
编号	字段中文名	字段英文缩写	字段英文全名	字段类型	字段长度
1	账户期初余额	ACCT_BEG_BAL	ACCOUNT BEGIN BALANCE	BIGINT	1
2	账户编码	ACCT_ID	ACCOUNT ID	BIGINT	
3	账户名称	ACCT_NAME	ACCOUNT NAME	VARCHAR	256
4	账户实时余额	ACCT_RL_TM_BAL	ACCOUNT REAL TIME BALANCE	BIGINT	
5	账户状态	ACCT_STATE	ACCOUNT STATE	SMALLINT	
6	账户状态名称	ACCT_STATE_NAME	ACCOUNT STATE NAME	VARCHAR	32
7	账户类型	ACCT_TYPE	ACCOUNT TYPE	SMALLINT	
8	活动标识	CMPN_FLAG	CAMPAIGN FLAG	SMALLINT	
9	活动编号	CMPN_ID	CAMPAIGN ID	VARCHAR	64
10	激活时间	ACT_DATE	ACTIVE DATE	DATETIME	
11	激活标识	ACT_FLAG	ACTIVE FLAG	SMALLINT	
12	片区描述	AREA_DESC	AREA DESCRIPTION	VARCHAR	256
13	片区编码	AREA_ID	AREA ID	BIGINT	
14	片区名称	AREA_NAME	AREA NAME	VARCHAR	64
15	区域类型	AREA_TYPE	AREA TYPE	SMALLINT	
16	ARPU 值	ARPU	AVERAGE REVENUE PER USER	BIGINT	
17	属性取值	ATTR_VAL	ATTRIBUTE VALUE	VARCHAR	64
18	报损时间	BAD_DEBT_DATE	BAD DEBT DATE	DATETIME	
19	呆坏账金额	BAD_DEBT_FEE	BAD DEBT FEE	BIGINT	
20	报损标识	BAD_DEBT_FLAG	BAD DEBT FLAG	SMALLINT	
21	保证金	BAIL_FEE	BAIL FEE	BIGINT	
22	开始日期	BEG_DATE	BEGIN DATE	DATETIME	
23	归属部门	BELO_DEPT_CODE	BELONG DEPARTMENT CODE	VARCHAR	16
24	大客户标识	BIG_CUST_FLAG	BIG CUSTOMER FLAG	SMALLINT	
25	账期	BILL_CYCLE	BILL CYCLE	VARCHAR	8
26	计费日期	BILL_DATE	BILL DATE	VARCHAR	8
27	账单金额	BILL_FEE	BILL FEE	BIGINT	
28	计费流量	BILL_FLUX	BILL FLUX	BIGINT	
29	手机号码	BILL_NO	BILL NO	VARCHAR	32
30	手机号段	BILL_NO_HEAD	BILL NO HEAD	VARCHAR	16
31	账单状态	BILL_STATE	BILL STATE	SMALLINT	
32	账单状态名称	BILL_STATE_NAME	BILL STATE NAME	VARCHAR	32
33	账单类型	BILL_TYPE	BILL TYPE	SMALLINT	
34	出生日期	BIRTH_DATE	BIRTH DATE	VARCHAR	8
35	黑名单标识	BLACK_LIST_FLAG	BLACK LIST FLAG	SMALLINT	

2、主数据和参考数据标准

主数据是用来描述企业核心业务实体的数据，比如客户、供应商、员工、产品、物料等；它是具有高业务价值的、可以在企业内跨越各个业务部门被重复使用的数据，被誉为企业的“黄金数据”。

参考数据是用于将其他数据进行分类或目录整编的数据，是规定数据元的域值范围。参照数据一般是有国标可以

参照的，固定不变的，或者是用于企业内部数据分类的，基本固定不变的数据。主数据与参照数据的标准化是企业数据标准化的核心。



3、 指标数据标准

指标类数据是指具备统计意义的基础类数据，通常由一个或以上的基础数据根据一定的统计规则计算而得到。

指标类数据标准一般分为基础指标标准和计算指标（又称组合指标）标准。基础指标具有特定业务和经济含义，且仅能通过基础类数据加工获得，计算指标通常由两个以上基础指标计算得出。

以下是某银行为确保指标数据标准定义的完整与严谨，形成的一整套指标数据标准的信息项属性架构：

基础属性	指标编号	指标大类	指标小类	指标细类	指标名称	指标别名	指标英文名	
业务属性	指标业务含义	指标业务口径	指标类型	指标制定依据	指标度量单位	是否报送监管	是否行内重点	统计频度
技术属性	指标取数范围	指标取数方式	指标条件	指标数据类型	指标数据长度	指标数据精度	指标敏感度	
管理属性	版本号	版本日期	定义部门	使用部门				

以“拨备覆盖率”指标为例，从数据标准化的角度来看，首先需要定义其业务含义，以明确其定位和用途，统一业务解释；同时通过技术属性明确其指标技术口径和取数规则等，确保指标数据计算结果的一致性。这样，在整个银行层面，统一了“拨备覆盖率”的业务口径和技术口径，最终确立了其使用规范。

基本属性	业务属性		技术属性				管理属性	
指标名称	指标业务含义	指标业务口径	数据类型	数据长度	数据精度	指标取数方式	定义部门	版本号
拨备覆盖率	贷款损失准备对不良贷款的比率。	贷款拨备率/不良贷款率 ×100%	比率类	8	4			

指标数据标准可以从 维度、 规则和 基础指标三个方面进行定义：



并非所有**模型数据**、**主数据**和**参考数据**、**指标数据**都应纳入数据标准的管辖范围。数据标准管辖的数据，通常只需要在各业务条线、各信息系统之间实现共享和交换的数据，以及为满足监控机构、上级主管部门、各级政府部门的数据报送要求而需要的数据。

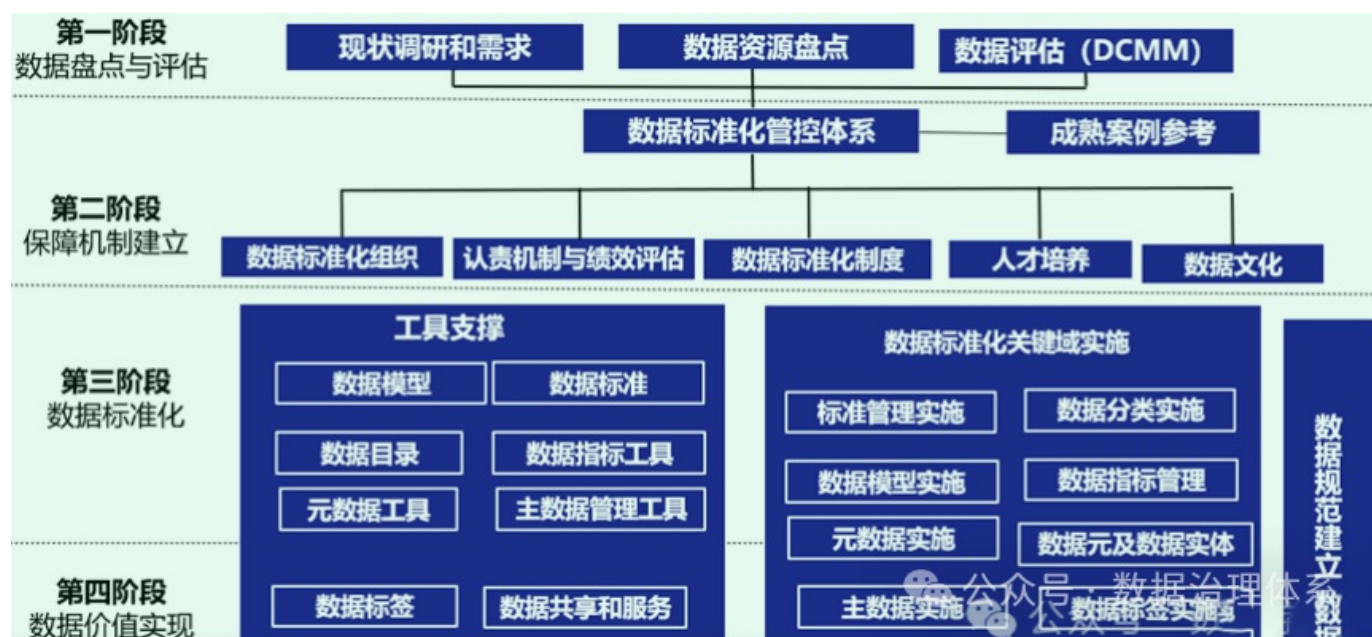
《数据标准管理实践白皮书》将数据分为**基础类数据**和**指标类数据**，数据标准也可以分为**基础类数据标准**或**指标类数据标准**，这种划分方法中的基础类数据其实等于模型数据+主数据+参考数据，因此不存在本质的区别。

三、数据标准化实施流程与方法

数据标准管理不但要解决好标准的制定和发布问题，更要解决标准如何落地（这是更重要的，也是更困难的）。企业做好标准体系规划、完成了各项标准的制定只是实现了数据标准管理的第一步，持续地贯标、落标，真正将各项标准应用于数据管理实践并充分发挥作用才是决定成功的关键。为了确保有效落标，除了组织推动、强化管理，还应遵循科学的工作流程，采用合理的技术手段和技术工具，真正实现科学落标、技术落标。

在构建数据标准化体系过程中，既要做好组织、人员和制度流程方面的准备，又要做好统筹规划和整体实施方案设计，确定好总体目标、阶段目标和实施路线图。同时企业要采取科学合理的实施方法，并配备相应技术平台及工具，持续、有效地推进各个阶段的工作任务直至目标达成。

企业数据标准化实施一般分四个阶段，如下图所示。



企业数据标准化实施阶段

第一阶段：数据盘点与评估

数据盘点与评估工作包括以下内容：

- 通过现状调研和需求收集充分了解数据标准化的现状和诉求；
- 通过数据资源盘点理清存量数据标准化的状况和问题；
- 结合数据评估和对标成功实践确定数据标准化的基线和目标，为数据标准化实施提供依据。

第二阶段：建立保障机制

建立保障机制工作包括以下内容：

- 建立数据标准化组织体系和认责流程，为数据标准化提供组织保障；
- 建立数据标准化制度规范，为数据标准化提供制度支撑；
- 加强人才培养和配置，为数据标准化提供人资和能力保障；
- 推进数据文化建设，为数据标准化营造良好的环境和氛围。

第三阶段：数据标准化实施

数据标准化实施包括以下内容：

- 建设技术平台和工具，为数据标准化实施落地提供技术支撑；
- 针对数据标准化各个关键域逐一开展专项实施，确保各项关键目标落地。

第四阶段：数据价值实现

数据标准化是企业数据治理的关键基础保障，也是数据治理的重要抓手，利用数据标准化驱动好企业数据治理，持续推动数据治理的成果落地、实现数据价值，才是企业开展数据标准化工作的终极目的。

四、数据标准如何落标？

数据标准的落标需要重点考虑三大问题：

问题1. 什么数据需要制定哪些标准

问题2. 什么系统落什么标准

问题3. 什么人与什么时间执行

如果这三个问题没有想清楚，基本数据标准的梳理会停留在Excel层面，标准的政策会停留在墙上，无法走入每个设计者的头脑和每个系统的每个字段。

第一个问题，什么数据需要制定标准，首先回到数据标准所要解决问题的初衷，数据标准主要解决数据在共享，融合，汇集应用中的不一致问题。那么看哪些数据会出现在这个这三个环节中，以及哪些容易出现问题。

对于与一个企事业组织来说，按照价值链，一般关注三大要素：客户，产品，大运营。举例来说，将银行业划分为九大概念数据，也是围绕客户与产品的大运营活动细分。

那么有如下几类数据会在数据应用过程中，会更多出现融合和汇总的机会，需要格外注意。

基础通用型数据	国家通用标准，行业通用标准，企业基础标准等
主数据类数据	客户,产品,渠道等主要经营数据
类型和维度类数据	分类码，标准码，维度码等
报送类数据	报送类模型和数据

第二个问题和第三个问题是实际工作中非常困扰的，落标的大多数困难与此有关，因此将其放在一起来说明，一般将系统与数据分列如下列表：

系统	落标强度	剖析	负责人	时机选择
核心业务系统数据	强	核心系统存储了企业经营最核心的数据，数据标准可以从核心系统定义中梳理，核心系统的数据的标准化会直接影响下游系统，数据流动链很长而系统改动难。因此尽量促成“我就是标准”这件事	开发团队 或 套装开发商	1.建设期强落标 2.维护期谨慎落标
重要业务系统数据	中	重要系统存储了核心系统的补充数据，重点是标准要与核心对齐	开发团队 或 套装开发商	1.建设期强落标 2.维护期谨慎落标
一般系统数据	弱	一般系统适合采用拉通型落标，重点在一些核心数据引用和类型类数据	开发团队 或 套装开发商	1.建设期选择重点落标 2.维护期可以不落标
报送和对外共享数据	强	遵从报送标准，和共享标准	报送团队	遵从报送和共享标准
DW初始汇集层	强	一般企业落标最容易着手之处，数据需要进行模型建模，汇集落标，需要进行ETL数据处理	数仓团队	尽早进行管控型强落标
DW汇总层数据	强	如果没有初始汇集层，可以在此针对具体主题进行落标，需要进行ETL数据处理。	数仓团队	全局规划后，应用开发时
Mart集市与报表数据	强	主要参照落标指标体系和维度体系，汇集落标，需要进行ETL数据处理	报表团队	每一次开发时

通过这个表格的内容，可以发现数据标准从源头落地，会减少数据的处理成本，提高数据应用的效益，缺点是对存量系统和外购系统存在较大改动风险和成本。

如果从数据的仓库层进行落标，比较容易着手处理，落标后的下游数据系统则自动统一数据标准，然而数仓层的报表应用与业务系统的报表存在口径不一致性在所难免，仍然需要源数据层进行必要调整。无论从哪一层入手，模型的优良设计环节都是必要条件，否则整个落标过程会没有抓手，流程将不顺畅。

国内某银行建立了一套数据标准体系框架，管理全行数据标准，形成了以科技战略委员会领导下的数据标准化小组为管理组织，涵盖标准定义、执行、监督评审等各个环节的良性数据标准闭环工作机制与流程，并配套管理制度和工具的建设。



其中包括建立基础数据标准600余条，覆盖了公用信息、产品、协议、资产、事件、渠道、参与人、财务8个主题的核心数据，建立代码数据标准200余条，将数据标准实施落地，保障核心数据的规范性和一致性。该系统的建设显著提升了向监管报送的数据质量，大大减少了IT部门数据质量处理任务工单，提升了取数效率。