



西安电子科技大学
XIDIAN UNIVERSITY

第七章 数据标准化

马 晶

经济与管理学院 信息管理系

Email: majing@xidian.edu.cn

课前讨论：标准是什么？

ICS 35.240.70
L 70



中华人民共和国国家标准

GB/T 38667—2020

信息技术 大数据 数据分类指南

Information technology—Big data—Guide for data classification



什么是“客户”？

财务会计说：“客户就是与我们发生业务往来、需要进行应收应付记账的组织或个人。”

销售人员说：“买了我们产品或服务的人就是我们的客户呀。”

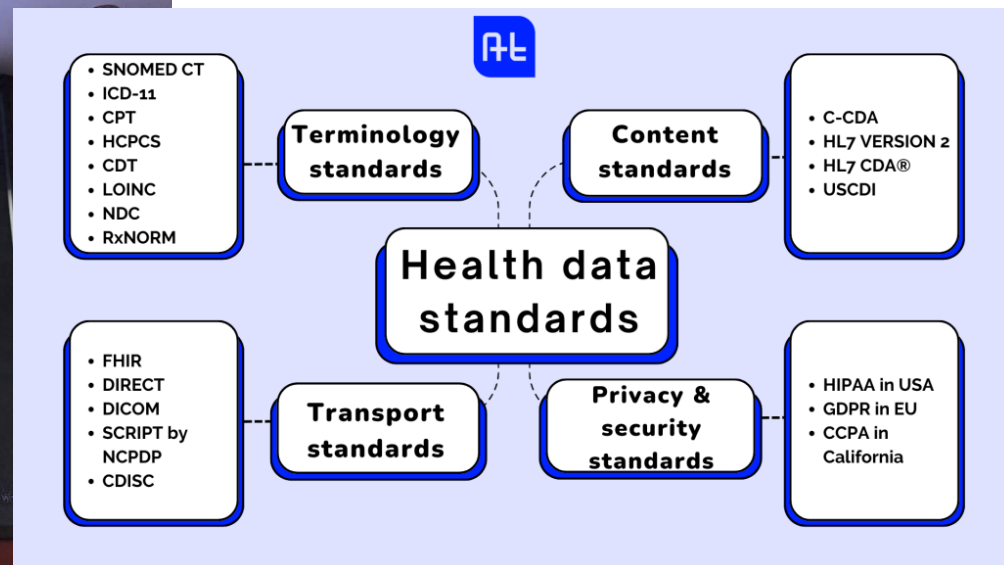
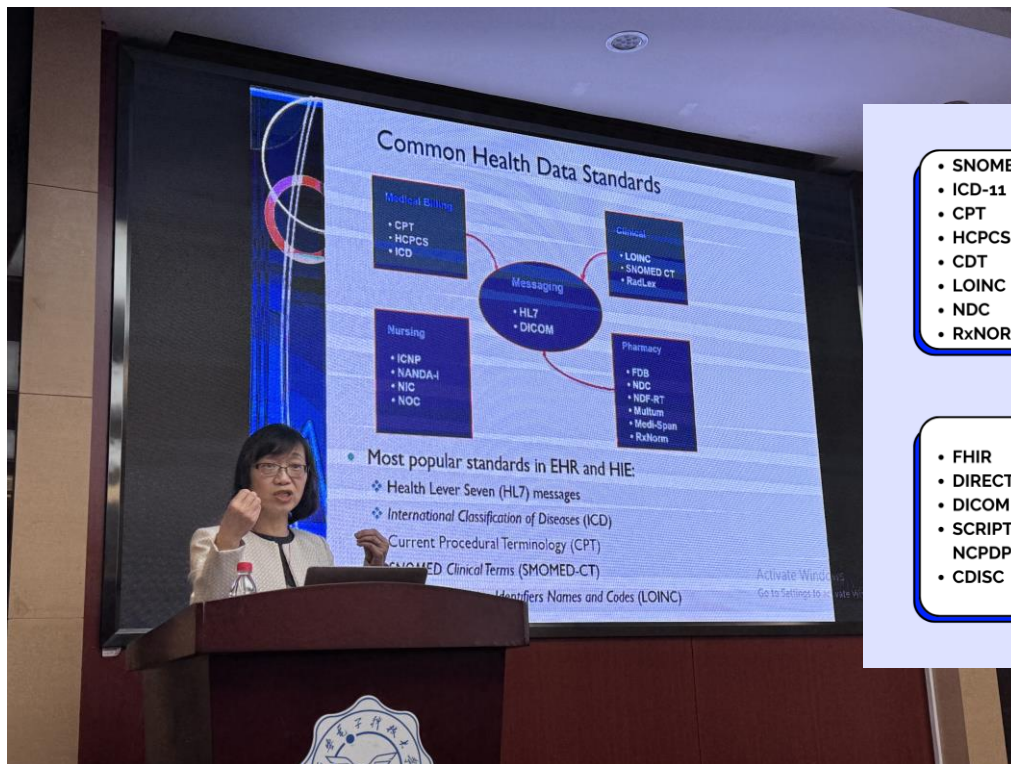
市场人员说：“有潜力和意向采购我们产品和服务的人都是我们的客户。”

。 。 。

所以，“客户”的标准是什么？

Healthcare Information Architecture and Data Modeling with Big Data

洪漪 2024.3.31



第七章 数据标准化



1. 眼科..... 1	7. 消化内科..... 101
1.1 疾病诊断名词..... 1	7.1 疾病诊断名词..... 101
1.2 症状体征名词..... 18	7.2 症状体征名词..... 116
1.3 手术操作名词..... 19	7.3 手术操作名词..... 117
1.4 临床检查名词..... 22	7.4 临床检查名词..... 118
2. 耳鼻喉科..... 24	8. 神经内科..... 119
2.1 疾病诊断名词..... 24	8.1 疾病诊断名词..... 119
2.2 症状体征名词..... 33	8.2 症状体征名词..... 144
2.3 手术操作名词..... 35	8.3 手术操作名词..... 145
2.4 临床检查名词..... 40	8.4 临床检查名词..... 145
3. 口腔科..... 42	9. 肾内科..... 150
3.1 疾病诊断名词..... 42	9.1 疾病诊断名词..... 150
3.2 症状体征名词..... 59	9.2 症状体征名词..... 155
3.3 手术操作名词..... 60	9.3 手术操作名词..... 155
3.4 临床检查名词..... 67	9.4 临床检查名词..... 156
4. 急诊科..... 69	10. 内分泌科..... 157
4.1 疾病诊断名词..... 69	10.1 疾病诊断名词..... 157
4.2 手术操作名词..... 70	10.2 症状体征名词..... 171
4.3 临床检查名词..... 71	10.3 手术操作名词..... 171
5. 心内科..... 72	10.4 临床检查名词..... 171
5.1 疾病诊断名词..... 72	11. 血液科..... 175
5.2 症状体征名词..... 79	11.1 疾病诊断名词..... 175
5.3 手术操作名词..... 80	11.2 症状体征名词..... 181
5.4 临床检查名词..... 81	11.3 手术操作名词..... 182
6. 呼吸内科..... 82	11.4 临床检查名词..... 182
6.1 疾病诊断名词..... 82	12. 普通外科..... 184
6.2 症状体征名词..... 92	12.1 疾病诊断名词..... 184
6.3 手术操作名词..... 93	12.2 症状体征名词..... 187
6.4 临床检查名词..... 96	

国家卫生健康委印发：常用临床医学名词(2023年版)

“由于人体解剖部位很多，疾病名称也多，以神经外科为例，解剖部位涉及额、颞、顶、枕、小脑、桥脑、中脑、延髓等，疾病名涉及脑膜瘤、胶质瘤、神经鞘瘤、室管膜瘤等，部位+病名有多个排列组合，

《常用临床医学名词》(2018年版)及本次增补词汇已将常见疾病尽可能收入，但仍然不能穷尽所有部位+病名的命名方法，故对未收录的名词可以按照“(解剖部位)+(疾病诊断)”的格式进行书写。”

什么是数据标准

- “无规矩，不成方圆”。数据标准管理就是为企业的数字化环境“建章立制”的过程。
- **数据标准 (Data Standard)** 是保障数据内外部使用一致、准确的规范和约束。
- 直观来说，数据标准可以理解为若干描述规范和要求的文档。
- 更严谨地，数据标准是为了使组织内部和外部使用交换的数据保持准确并且一致，经协商一致制定的并由相关主管机构批准，共同使用和重复使用的一种**规范性文件**。
- 更一般地，数据标准不仅仅是一套规范，更是由**管理规范、管控流程、技术工具**共同组成的**体系**，使通过这套体系逐步实现信息标准化的过程。

- 数据标准的分类和数据的分类是对应的，下面给出几种典型的定义和划分方法：

1. 根据数据类型

- 一种典型的方式是将数据划分为基础类和指标类，并进一步定义和规范相应的数据标准。
- ◆ **基础类数据**：是指企业或组织在工作和生产的过程中产生的、可直接使用、未经统计分析的基本数据，如商品价格、商品类型、商品编号等。基础类数据标准主要是为了保证技术类数据使用的一致准确，其约束了业务活动产生的数据的类型、值域、标识符等规范。
- ◆ **指标类数据**：是指经过统计分析，具备统计意义和指标作用的数据，如均值、增幅、中位数等。指标类标准又可以进一步分为基础指标标准和计算指标标准，分别针对基础数据上直接计算得到的指标和不同指标二次合成得到的指标。

例：每股收益

名称：每股收益

英文：Earnings Per Share

简称：EPS

同义词：每股盈余，每股盈利，每股税后利润

定义：企业税后利润与股本总数的比率。是普通股股东每持有一股所能享有的企业净利润或需承担的企业净亏损。

计算公式：基本每股收益=归属于普通股股东的当期净利润/当期发行在外普通股的加权平均数

应用：每股盈余通常被用来反映企业的经营成果，衡量普通股的获利水平及投资风险，是投资者等信息使用者据以评价企业盈利能力、预测企业成长潜力、进而做出相关经济决策的重要的财务指标之一。

数据关联：利润表，资产负债表

执行标准：财政部令《企业会计准则》第34号

数据类型：Double (8)

精度：小数点后两位

单位：元/股

2. 根据数据处理和使用阶段

可将数据标准定义和划分为数据处理标准、数据安全标准、数据质量标准、产品和平台标准、应用和服务标准等不同类别。

- ◆ **数据处理标准**包括针对数据整理、数据分析、数据访问的标准。
- ◆ **数据安全标准**包括通用标准和隐私保护标准。
- ◆ **数据质量标准**包括元数据质量、质量评价、数据溯源等方面的标准。
- ◆ **产品和平台标准**包括关系型数据库产品、非结构化数据管理产品、可视化工具、数据处理平台方面的标准。
- ◆ **应用和服务标准**包括开放数据集、数据服务平台等方面的标准。

3. 根据数据标准涉及的实体

根据涉及的实体，可以定义数据模型和数据规范。

- ◆ **数据模型**包括数据模型本身和业务规则的定义。
- ◆ **数据规范**包括数据建模规范、数据编码规范、数据交换规范、数据ETL规范、数据集
成规范等。

数据标准的作用

数据标准适用于业务数据描述、信息管理及应用系统开发，既可作为经营管理中所涉及数据的规范化定义和统一解释，也可作为数据管理的基础，同时也是在应用系统开发时进行数据定义的依据。

示例：数据标准在数据使用中的作用

在一个数据库中，客户数据模型允许存储三个地址：**交货地址、账单地址和备用联系人地址**。前两个字段填充了实际地址，但第三个字段一直为空。由于某些业务需要，一段时间后第三个字段成为维护客户的注释信息，例如“免税”“现金支付”等。实际上，仔细检查就会发现，在名称字段中也嵌入了类似的注释。

采用这种使用模式，久而久之就造成了数据的重复、不完整、不准确等诸多问题。

一种解决方案是**清理数据**，即直接对名称、地址等字段进行清理，删除多余的数据。但在这种方式下，清理客户主数据的名称和地址会给业务人员带来困扰，因为有时候它们是有意义的。

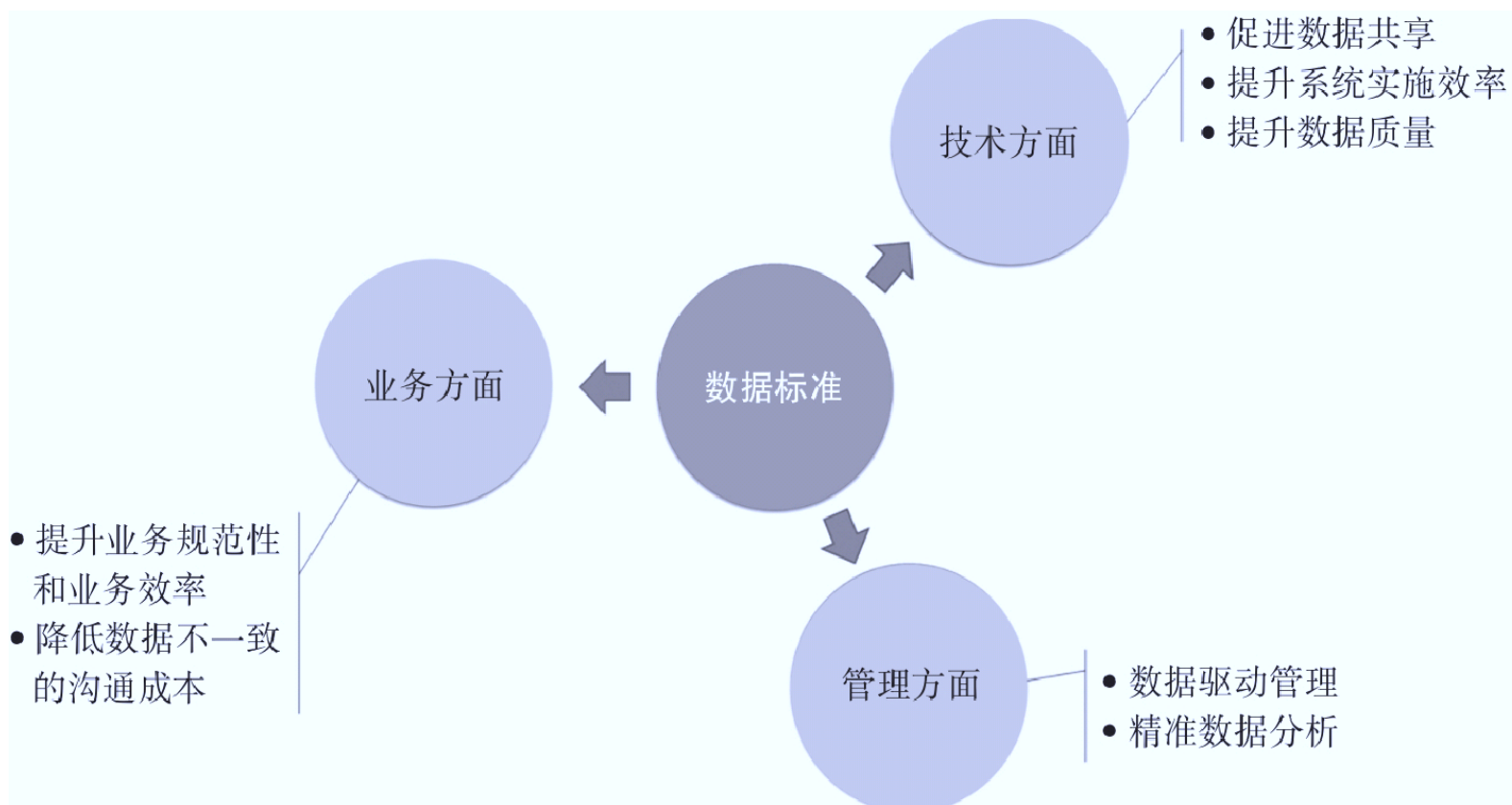
因此在创建“客户”这个数据模型时，需要**对模型涉及的每个字段进行定义，明确每个数据属性的语义、用途、结构、业务规则及填写规范**。这就是数据标准的用武之地，这将有助于防止数据对象、数据属性的定义之间的冲突。

数据标准的作用

在企业的数据管理和数据应用中，数据标准除了能防止数据对象、数据属性的定义之间的冲突，还对企业应用系统的集成和数据分析挖掘具有重要意义。

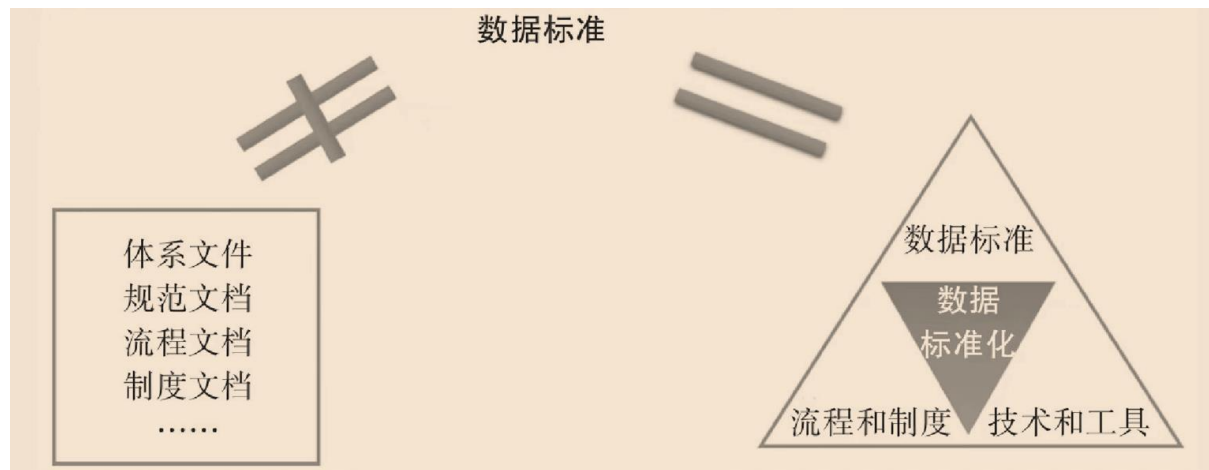
- 数据标准可以增强各业务部门对数据理解的一致性，提升沟通效率。
- 数据标准可以减少数据转换，促进系统集成和信息资源的共享。
- 数据标准可以促进企业级单一数据视图的形成，支持数据管理能力的发展。
- 数据标准有助于对数据进行统一规范的管理，消除各部门间的数据壁垒，支持业务流程的规范化。
- 数据标准有利于提高数据质量。可以基于数据标准的规范化定义对企业数据质量进行检查，找出有问题的数据，出具数据质检报告。
- 数据标准有利于规范化管理数据资产。数据标准是数据资产梳理和定义的基础。对于一家拥有大量数据资产或者要实现数据资产交易的企业而言，构建数据标准是一件必须做的事情。

数据标准管理的意义



数据标准化

- 严格来讲，将数据的分类分级、记录格式及转换、编码以及与其相关的管理制度、流程等编制成文件并不等于完整的数据标准化。企业数据的标准化更多的是一组涉及数据标准制定、数据标准管理流程和制度、数据标准管理技术和工具的解决方案。
- (1) 数据标准化是建立各部门共识的过程。
 - (2) 数据标准是各系统数据整合的基础。



数据标准管理的阶段

数据标准管理是指研究、制定、应用和维护数据标准的过程和活动。基于这个定义，可以大致分为4个主要阶段来讨论数据标准管理。

1. 标准研究

研究阶段主要的目的是回答两个问题，即“标准是否需要”和“需要何种标准”。工作内容主要包括**标准规划**和**标准需求管理**。

◆ **标准规划**：包括标准调研、案例收集、业务分析、制订计划等活动。标准调研和案例收集的主要目的是调研现有标准并收集相关领域的最佳实践，业务分析和制定计划则根据本组织或企业，了解与标准相关的内容，确定所需标准的大致范围和框架。

◆ **标准需求管理**：指定专门的人员组织、收集、精练具体的标准，并根据规范流程和语言来描述标准需求。

2. 标准制定

在标准制定阶段，依据前期的调研和需求分析结果，参考国际及国家标准化组织的相关标准，完成标准的细节定义。具体包括：统一的数据定义标准和编码标准；清晰定义每个数据元素的名称、内容、类别、关联关系、用途、源系统等属性；保证可行性、可扩展性、可共享性、稳定性等数据标准应当具备的基本性质。

3. 标准应用

标准应用的主要目的是将制定好的标准实施到各个部门和系统中。对于新开发的系统，要遵循数据标准中给出的数据编码规范、数据接口规范、数据建模规范、数据交换规范等；对于已存在的系统，如不符合数据标准，则制定相应计划，分阶段、分批次的完成修正，直至满足标准。

4. 标准维护

在数据标准维护阶段，需要有相应的需求收集、需求评审、变更评审、发布等多个步骤，并能对所有的修订做版本管理，以方便将来问题查找。

数据标准管理的常见问题

(1) 数据语义不清晰

当独立使用一个系统时，相关业务术语、相关联语义可能是一致的，但如果需要在两个或多个环境之间比较值，含义上的细微差别就会被放大。

(2) 数据定义和使用语境

数据定义的歧义主要表现在同名异义、同义异名的情况。

同名异义是指名称相同但代表的含义不同，常见的是相同名称的数据在不同的语境中所代表的含义是不同的；同义异名是指含义相同但命名不同的情况，比如，同样的“姓名”有“员工姓名”和“职工姓名”两种叫法，很可能开发人员给它们定义的标识分别为“YGXM”和“ZGXM”。

(3) 标准的制定和使用两层皮

数据标准是数据一致性、完整性、准确性的保证，是数据分析、数据集成的基础。数据标准的建立需要经过审批、发布，再在被治理系统中进行推广和使用。同时，还需要评估数据标准的落地情况，通过评估定位数据问题并进行整改，以保证制定出的数据标准被正确使用，避免标准制定和标准执行两层皮。

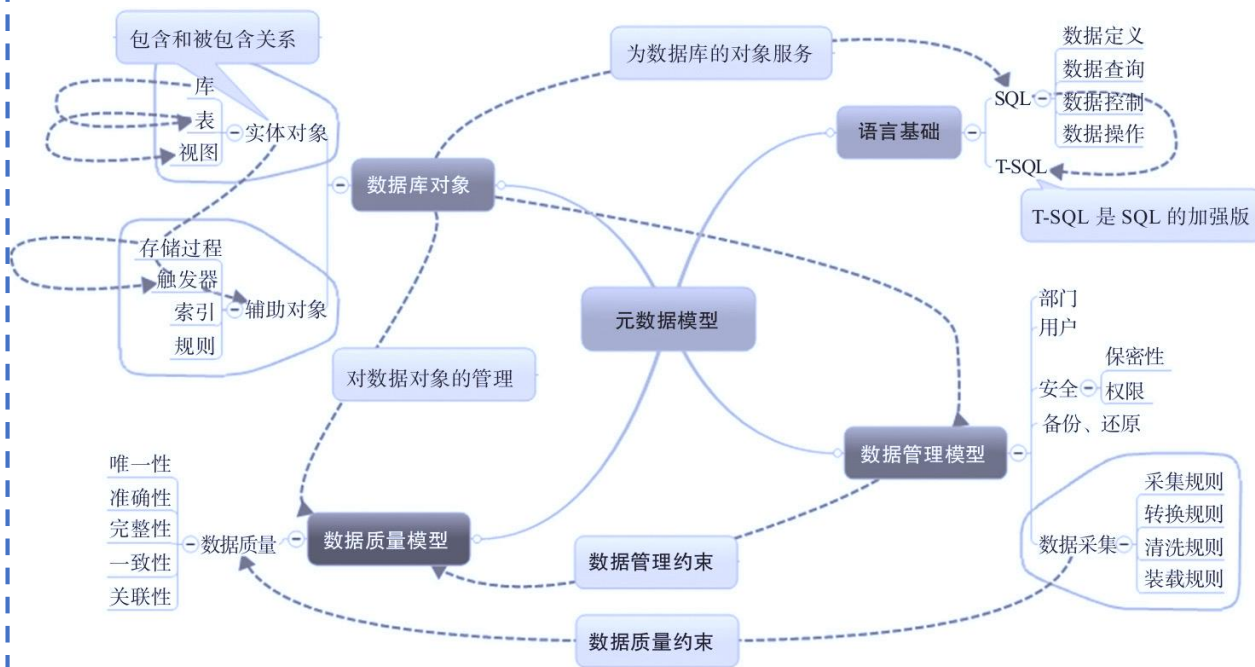
数据标准管理内容

企业的数据标准一般包含4方面内容：**数据模型标准、基础数据标准、主数据与参考数据标准、指标数据标准。**

1. 数据模型标准

数据模型反映的是对业务的理解和定义，能够帮助企业建立组织内部和组织之间沟通的桥梁。

数据模型标准通过技术元数据、业务元数据进行模型描述，将业务信息和技术信息完整体现在数据模型中，并确保数据模型能够准确、完整地反映业务需求和相关技术约束，要求规范化、标准性、一致性、可读性。

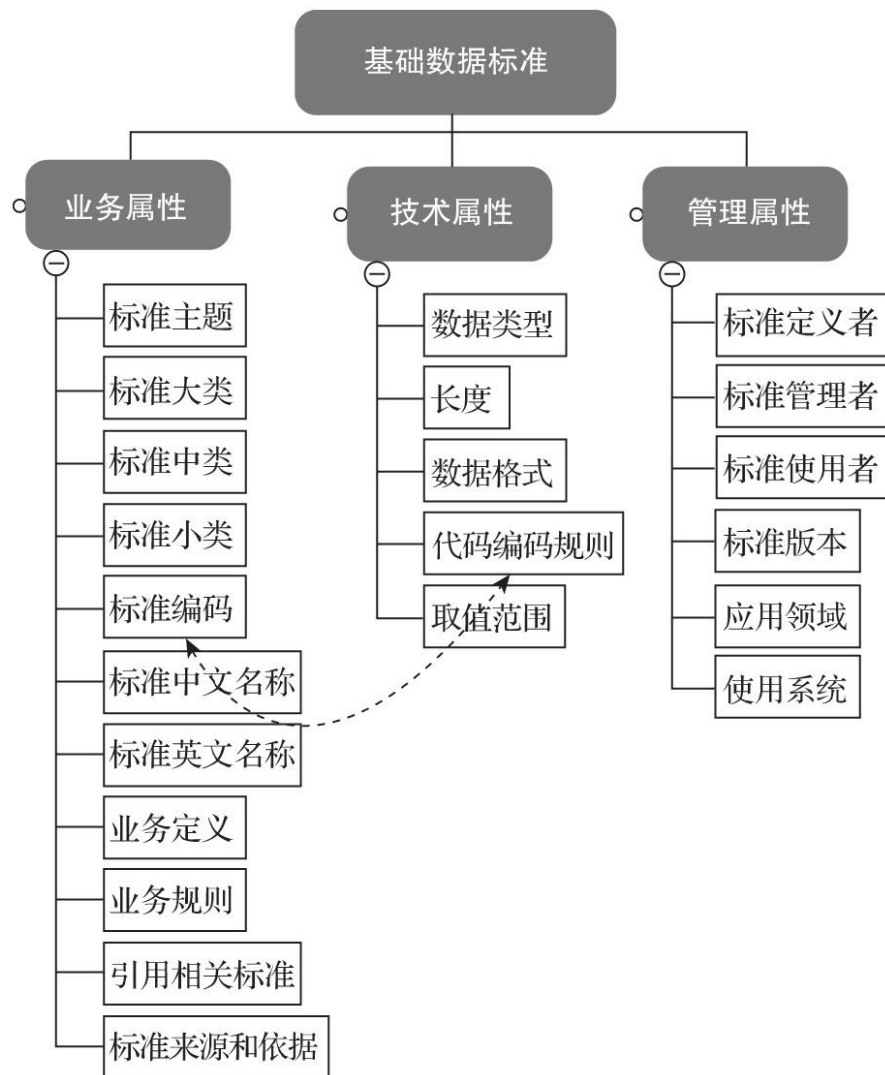


2. 基础数据标准

基础数据是系统的数据字典，在系统初始化时就存在于系统数据库中，是结构性或功能性的支撑，如国家地区、行政区划、邮政编码、性别代码、计量单位代码等。

基础数据标准一般会涉及国际标准、国家标准和行业标准。在定义数据实体或元素时可以引用相关标准，再根据企业的需求不断补充完善、更新优化和积累，以便更好地支撑业务应用的开发、信息系统的集成和企业数据的管理。

基础数据标准通常用来对应用系统或数据仓库的数据字典进行标准化，一般包含业务属性、技术属性、管理属性三部分。

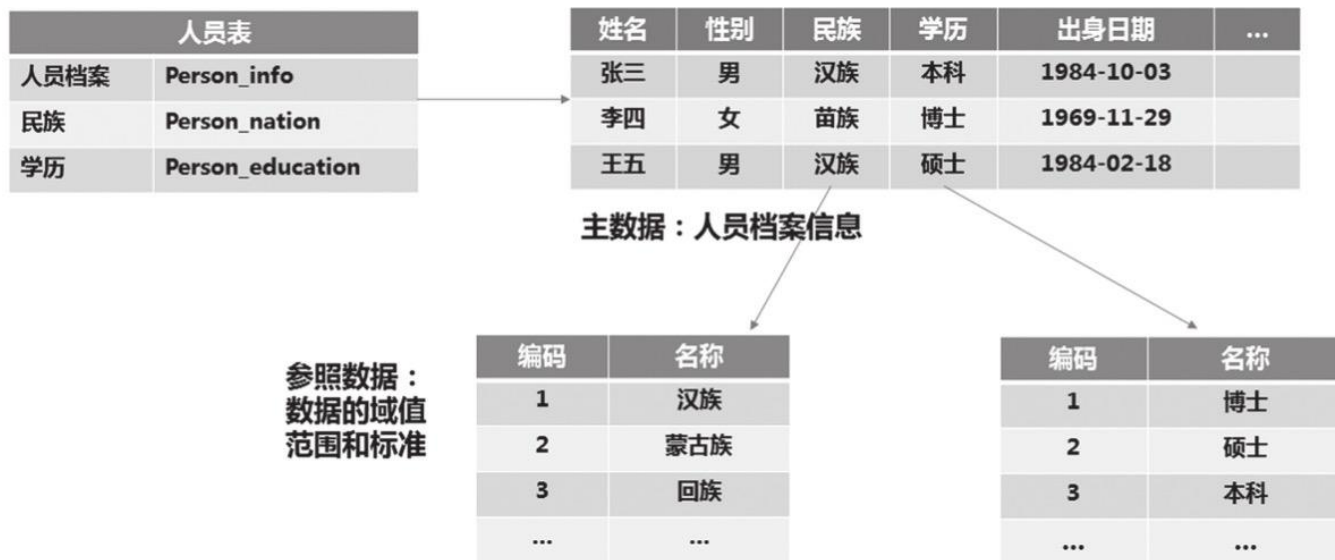


标准编号	CD190004	
中文名称	性别代码	
英文名称	Codes for sexual distinction of human	
代码描述	描述人的性别代码	
定义原则	采用外部标准	
引用标准代号及名称	GB/T 2261.1—2003 个人基本信息分类与代码 第 1 部分：人的性别代码	
代码编码规则	1 级 2 位编码（1，2），采用国标编码	
技术属性	CHAR（2）	
版本日期	2019/11/18	
标准类别	标准	
代码值	代码描述	业务说明
01	男性	
02	女性	
99	未说明的性别	

3. 主数据与参考数据标准

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、可以在企业内跨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，是规定数据属性的域值范围。参考数据一般以国际标准、国家标准或行业标准为依据，是固定不变的数据。



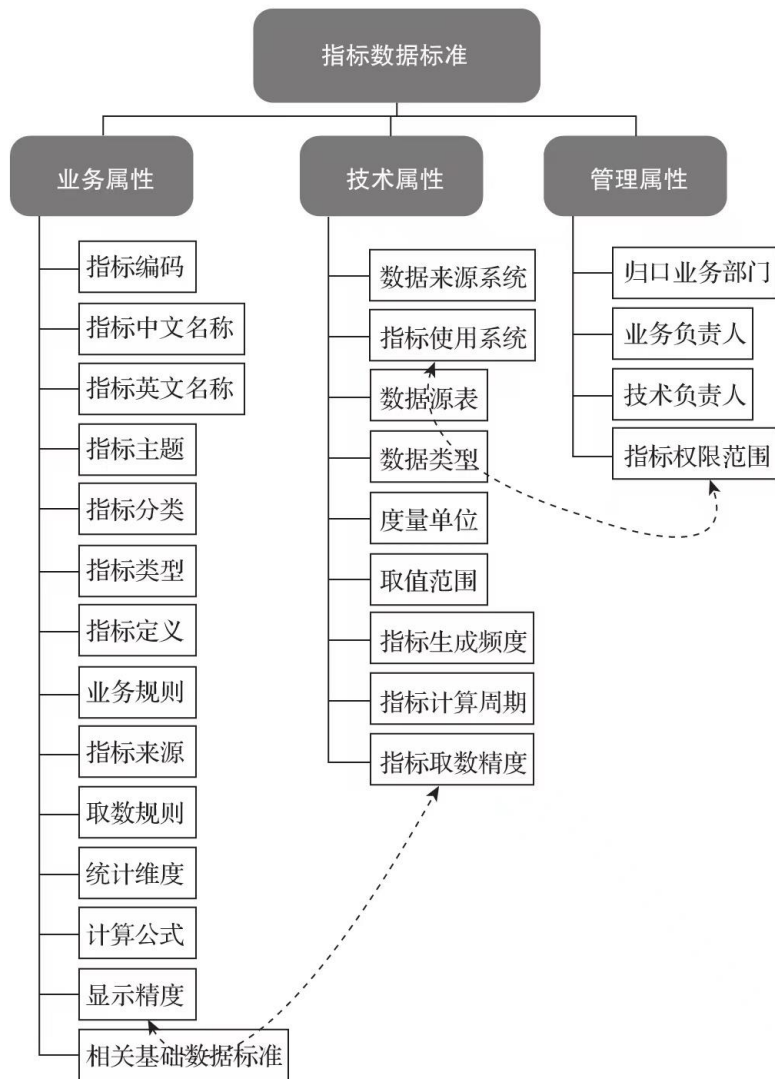
- 主数据标准包含主数据分类、主数据编码和主数据模型。
- 主数据分类是根据主数据的属性或特征，将其按一定的原则和方法进行区分和归类，并建立起一定的分类体系和排列顺序。
- 主数据编码是为事物或概念（编码对象）赋予具有一定规律、易于计算机和人识别处理的符号，形成代码元素集合。对各类主数据概念的正确理解依赖于主数据分类，对各类主数据作出唯一表示依赖于主数据编码。
- 主数据模型即基于主数据属性的逻辑模型或物理模型，包括每个属性的名称、属性性质、类型、质量规则、取值范围等。

主数据	人员			
定义	人员主数据是指所有与企业签署了正式劳动合同的人员，人员主数据是从企业管理视角出发的人员实体的数字化描述			
序号	属性名称	属性性质	类型	取值范围
1	人员编码	系统自动生成	字符型	系统自动生成的 7 位流水码
2	姓名	必填项	字符型	集团员工姓名，同身份证上的名称一致，必须保证姓名输入准确
3	身份证号	必填项	字符型	位数为 15 位或 18 位的身份证号码（港澳台及外籍除外）
4	性别	必填项	枚举型	男；女
5	出生日期	必填项	日期型	须与身份证上的出生日期保持一致
6	民族	必填项	参照型	参照民族档案
7	电子邮件	必填项	字符型	不能为空，格式：zhangsan@cn-nthq.com
8	办公电话		字符型	区号加 6~8 位电话号码（+分机号），中间以“-”连接，如 010-12345678 或 010-12345678-8888
9	手机	必填项	字符型	位数默认为 11 位（港澳台及国外除外），不得以其他符号代替
10	学历		枚举型	小学、初中、高中、大专、本科、硕士、博士
11	状态	必填项	枚举型	在职；离职
12	备注		字符型	

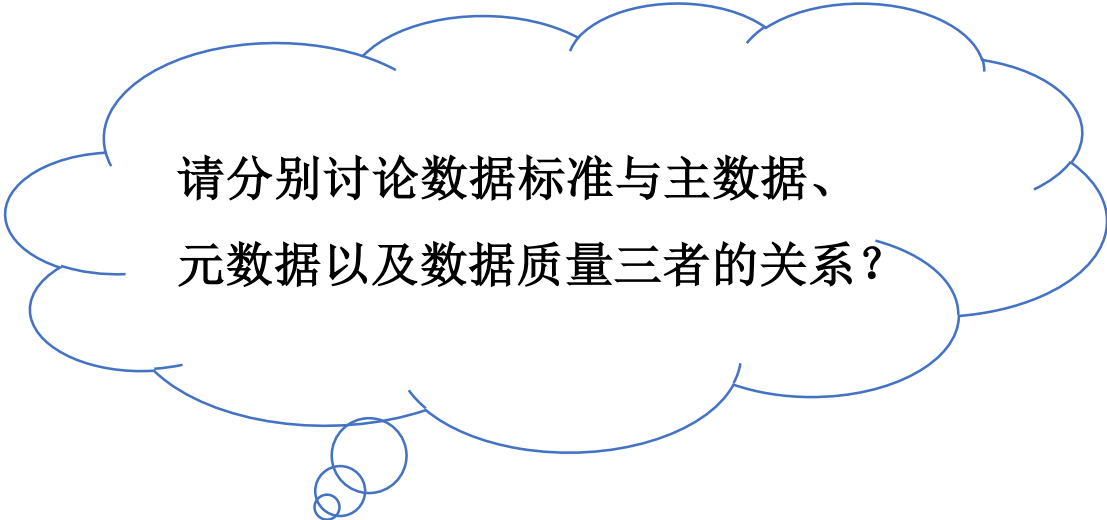
4. 指标数据标准

企业的各业务域、各部门均有其相应的**业务指标**，这些指标有的名称相同却有着不同的业务含义，而有的指标虽然名称差异很大，但在业务上却是同一个指标。如果不对指标数据进行标准化，你可以想象：对于同一指标，不同系统的指标统计结果可能是不同的，而且很难分清哪个才是正确的；每次有新分析主题构建或旧分析主题变更时，都需要从所涉及的各个系统、库表中重新定义指标，成本很高。另外，目前大数据分析都提倡业务人员自助分析，如果没有指标数据标准，业务人员几乎不可能从不同系统中拿到自己想要的数​​据，自助式分析将无从谈起。

指标数据标准是在基础类数据基础之上增加了**统计维度、计算方式、分析规则**等信息加工后的数据，它是对企业业务指标所涉及指标项的统一定义和管理。指标数据标准与基础数据标准一样，也包含业务属性、技术属性、管理属性三部分。



问题思考



请分别讨论数据标准与主数据、
元数据以及数据质量三者的关系？

（1） 数据标准与主数据的关系

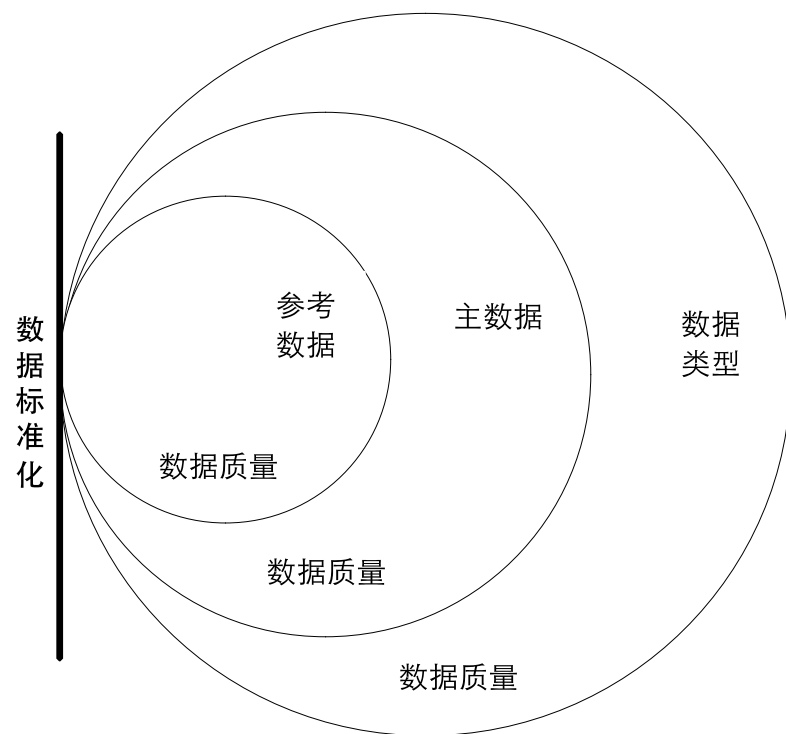
从范围上看，数据标准包含主数据与参考数据标准。主数据标准包括主数据分类、主数据编码和主数据模型等。

（2） 数据标准与元数据的关系

组织在制定数据标准的时候最先需要明确的就是数据实体的属性，包括属性的定义、业务规则、关系等，这些信息在元数据管理中叫做业务元数据。

（3） 数据标准与数据质量的关系

标准化是信息化的前提，是组织实施数据质量管理的基础。通过对数据标准的统一定义，明确数据的归口部门和责任主体，为组织的数据质量和数据安全提供基础保障。





本章要点

- 掌握数据标准与数据标准化的概念
- 熟悉数据标准管理阶段及管理内容