

考试资料职业发展
技术读书笔记分享

B站/闲鱼：大西洋活跃的锅巴
公众号：不太甜

DAMA-DMBOK

数据管理知识体系指南CDGA/CDGP认证

第10章 参考数据和主数据（完整课程视频请扫描二维码）



第10章 参考数据和主数据



01

引言

02

活动

03

工具和方法

04

实施指南

05

治理



01

引言

定义、业务驱动因素、目标和原则、基本概念

参考数据：

参考数据不易变化，它的数据集通常会比交易数据集或主数据集小、复杂度低，拥有的列和行也更少。

主数据：区别是什么

主数据是以与业务活动相关的通用和抽象概念形式提供业务活动语境的数据，包括业务交易中涉及的内部和外部对象的详细信息。

主数据面临的主要挑战是实体解析，也称身份管理，它是识别和管理来自不同系统和流程的数据之间的关联的过程。

- 1) 满足组织数据需求
- 2) 管理数据质量
- 3) 管理数据集成的成本
- 4) 降低风险

集中管理的参考数据会使组织获得如下好处：

- 1) 通过使用一致的参考数据，满足多个项目的数据需求，降低数据整合的风险和成本
- 2) 提升参考数据的质量

目标：

- 1) 确保组织在各个流程中都拥有完整、一致、最新且权威的参考数据和主数据
- 2) 促使企业在各业务单元和各应用系统之间共享参考数据和主数据
- 3) 通过采用标准的、通用的数据模型和整合模式，降低数据使用和数据整合的成本及其复杂性。

原则：

- 1) 共享数据
- 2) 所有权
- 3) 质量
- 4) 管理职责
- 5) 控制变更

①在给定的时间点，主数据值应该代表组织对准确和最新内容的最佳理解。改变数据值的匹配规则，应该在有关监督下谨慎地运用。任何合并或拆分参考数据和主数据的操作都应该是可追溯的。

②对参考数据的更改应该遵循一个明确的流程：在实施变更之前应该进行沟通并得到批准

6) 权限

1、主数据

奇泽姆六层数据分类法

元数据、参考数据、企业结构数据、交易结构数据、交易活动数据和交易审计数据。

主数据定义为参考数据、企业结构数据和交易结构数据的聚合。

(1) 参考数据

代码表和描述表，仅用于描述组织中的其他数据，或者仅用于将数据库中的数据与组织之外的信息联系起来。

(2) 企业结构数据

会计科目表

(3) 交易结构数据

客户标识符

参考数据和主数据的管理重点不同：考题

- 1) 参考数据管理（RDM）。需要对定义的域值及其定义进行控制。参考数据管理的目标是确保组织能够访问每个概念的一整套准确且最新的值。
- 2) 主数据管理（MDM）。需要对主数据的值和标识符进行控制，以便能够跨系统地、一致地使用核心业务实体中最准确、最及时的数据。目标包括确保当前值的准确性和可用性，同时降低由那些不明确的标识符所引发的相关风险。

是指可用于描述或分类其他数据，或者将数据与组织外部的信息联系起来的任何数据。

参考数据常用的存储技术包括：

- 1) 关系数据库中的代码表
- 2) 参考数据管理系统
- 3) 用特定于对象属性的元数据指定允许值，重点在于应用程序的调用接口或用户界面访问。

(1) 参考数据结构

- 1) 列表：由代码值和代码描述组成的列表。
- 2) 交叉参考数据列表：不同的应用程序可以使用不同的代码集表示相同的概念。
- 3) 分类法：根据不同级别的差异性获取信息；
利用内容分类和多方位的导航以支持商务智能；
也可以维护数据层次信息。
- 4) 本体：将用于管理网站内容的本体作为参考数据的一部分。

(2) 专有或内部参考数据

创建参考数据来支持内部流程和应用

(3) 行业参考数据

用于描述由行业协会或政府机构而不是由某个组织创建和维护的数据集

(4) 地理或地理统计参考数据

(5) 计算参考数据

如外汇依赖于及时更新的交换汇率值表

(6) 标准参考数据集的元数据

主数据是有关业务实体的数据，这些实体为业务交易和分析提供了语境信息。
一般组织的主数据包括下列事物的数据：

- 1) 参与方
- 2) 产品和服务
- 3) 财务体系
- 4) 位置信息

(1) 记录系统，参考系统

记录系统是一个权威的系统，它通过使用一套定义好的规则和预期来创建、获取并维护数据。

参考系统也是一个权威系统，数据消费者可以从参考系统中获得可靠的数据来支持交易和分析，即使这些信息并非起源于参考系统。

主数据管理应用（MDM）、数据共享中心（DSH）、数据仓库DW通常会被用作参考系统

(2) 可信来源，黄金记录

基于自动规则和数据内容的手动管理的结合，可信来源是事实的最佳版本；

可信来源也被称为单一视图、360度视图；

在可信来源中，表示一个实体、实例的最准确数据的记录可以被称为黄金记录。

（3）主数据管理

一个技术支持的领域，在这个过程中业务和技术协同工作，以确保企业官方共享主数据资产的统一性、准确性、管理性、语义一致性和问责性。主数据是由标识符和扩展属性组成的一个一致且统一的集合，它描述了企业的核心实体，包括客户、潜在客户、企业公民、供应商、位置、层次结构和会计科目等。

评估一个组织的主数据管理情况，需要识别以下几点：

- 1) 哪些角色、组织、地点和事物被反复引用
- 2) 哪些数据被用来描述人、组织、地点和事物。
- 3) 数据是如何被定义和设计的，以及数据粒度细化程度如何。
- 4) 数据在哪里被创建或来源于哪里，在哪里被储存、提供和访问
- 5) 数据通过组织内的系统时是如何变化的
- 6) 谁使用这些数据，为了什么目的
- 7) 用什么标准来衡量数据及其来源的质量和可靠性

(3) 主数据管理

在一个域内规划主数据管理包括以下几个基本步骤：

- 1) 识别能提供主数据实体全面视图的候选数据源
- 2) 为精确匹配和合并实体、实例制定规则
- 3) 建立识别和恢复未恰当匹配或合并的数据的方法
- 4) 建立将可信数据分发到整个企业的系统中的方法

(3) 主数据管理

全生命周期中的关键活动包括：

- 1) 建立主数据实体的上下文，包括相关属性的定义及其使用条件，并加以治理
- 2) 识别出在单个数据源内以及多个数据源中代表同一实体的多个实例；构建并维护标识符和交叉引用，以支持信息整合
- 3) 协调和整合不同来源的数据，以提供主记录或事实的最佳版本。
- 4) 识别出那些未被正确匹配或合并的实例，确保它们得到修正，并关联到正确的标识符
- 5) 通过直接存取、使用数据服务，或通过复制反馈到交易系统、数据仓库或其他分析性数据存储系统，实现对可信数据的跨程序访问
- 6) 在组织内强制使用主数据

(4) 主数据管理的关键处理步骤

包括数据模型管理、数据采集、数据验证、标准化和数据丰富、实体解析、管理和共享。

1) 数据模型管理

主数据工作揭示了清晰一致的逻辑数据定义的重要性。这些模型应该能够帮助组织克服“系统发言”的状况。

2) 数据采集

- ①接收并应对新的数据源采集的要求
- ②使用数据清理和数据分析工具进行快速、即时、匹配或高级的数据质量评估
- ③评估数据并将数据整合的复杂性传递给请求者，以帮助他们进行成本效益分析
- ④试点数据采集及其对匹配规则的影响
- ⑤为新数据源确定数据质量指标
- ⑥确定由谁负责监控和维护新数据源数据的质量
- ⑦完成与整体数据管理环境的集成

(4) 主数据管理的关键处理步骤

3) 数据验证、标准化和数据丰富

①验证

识别那些被证明是错误的或可能是不正确或默认的数据

②标准化

确保数据内容符合标准参考数据值、标准的格式或字段

③数据丰富

添加可以改进实体解析服务的属性。

(4) 主数据管理的关键处理步骤

4) 实体解析和标识符管理

实体解析时确定两个对现实世界对象的引用到底是指同一对象还是不同对象的过程。

实体解析包括一系列活动（实例提取、实例准备、实例解析、身份管理、关系分析），这些活动能够使实体、实例的身份以及实体、实例之间的关系持续地被管理。

①匹配

a) 假阳性：不代表同一实体的两个对象被关联在一个标识符下，假阳性导致一个标识符指向多个现实世界的实体实例

b) 假阴性：代表同一实体的两个对象没有被关联在一个标识符下。假阴性导致多个标识符指向同一现实世界的实体，但一个实体本应该只有一个标识符
相似性分析或匹配，通常基于相应属性值之间的加权近似匹配对两个记录之间的相似程度进行评分。

a) 确定式算法：解析和标准化依靠确定的模式和规则，按照指定的权重计算相似度的分值

b) 或然式算法：依赖于训练数据的采集能力，训练是需要观察全部记录的一个子集的预期结果，再进行匹配器调优，实现匹配器基于统计结果的自我调整。

(4) 主数据管理的关键处理步骤

4) 实体解析和标识符管理

②身份解析

③匹配流程/协调类型

a) 重复表示匹配规则：重点关注一组特定的数据元素，这些元素能够唯一地标识实体，识别合并机会而不采取自动合并操作。

b) 匹配链接规则：标识可能与主记录相关的记录，只建立交叉引用关系而不更新这些被交叉引用的记录的内容；更容易实现也容易撤销

c) 匹配合并规则：重点关注记录的匹配，并将这些记录中的数据合并为单一、统一、协调且全面的记录。

（4）主数据管理的关键处理步骤

4) 实体解析和标识符管理

④主数据ID管理

管理主数据涉及管理标识符，在主数据管理环境中，有两种类型的标识符需要跨数据源管理：**全局标识符（Global ID）、交叉引用（X-Ref）信息**

全局标识符是主数据管理解决方案分配和维护的唯一标识符，它会被附加到协调记录中，其目的是能够唯一地标识一个实体实例。

全局标识符（Global ID）都应该只由一个授权解决方案生成，以避免出现重复值的风险。全局标识符需要解决的难题是，在发生反合并/再合并的情况下如何维护正确的全局标识符。

交叉引用管理（X-Ref Management）应该包括维护此类映射历史的功能以支持匹配率指标，并开放查找服务以支持数据整合

⑤从属关系管理

- a) 隶属关系
- b) 父子关系
- c) 数据共享和管理责任

(5) 参与方主数据

参与方主数据是关于个人、组织以及他们在业务关系中所扮演的角色的数据。

面临的挑战：

- 1) 个人和组织扮演的角色和他们之间关系的复杂性
- 2) 唯一标识的困难
- 3) 数据源的数量和它们之间的差异
- 4) 多个移动通信信道和社交渠道
- 5) 数据的重要性
- 6) 客户想要怎样参与的期望

（6）财务主数据

包括有关业务部门、成本中心、利润中心、总账账户、预算、计划和项目的数据。ERP系统充当财务主数据的中心枢纽。

（7）法律主数据

包括关于合同、法规和其他法律事务的数据。允许对提供相同产品或服务不同实体的合同进行分析以便更好地协商谈判，或将这些合同合并到主协议中。

（8）产品主数据

专注于组织的内部产品和服务，或全行业的产品和服务。不同类型的产品主数据解决方案支持不同的业务功能。

1) 产品生命周期管理（PLM）系统侧重于从构想、开发、制造、销售、交付、服务和废弃等方面管理产品或服务的生命周期。组织通过实施产品生命周期管理系统以加快产品的上市。

2) 产品数据管理（PDM）系统通过捕获和实现对设计文档（如CAD图样）、配方（制造说明书）、标准操作程序和物料清单（BOM）等产品信息的安全共享，以支持工程和制造功能。产品数据管理功能可以通过专门的系统或ERP系统实现。

3) 企业资源规划（ERP）系统的产品数据主要关注库存单位，以支持从订单录入到库存阶段，可以通过多种技术识别各种独立的产品。

4) 制造执行系统（MES）中的产品数据主要关注原材料库存、半成品和成品，其中成品与可以通过ERP系统来存储和订购的产品相关联。

5) 客户关系管理（CRM）系统支持营销、销售和交互支持，系统中的产品数据可以包括产品系列和品牌、销售代表协会、客户区域管理以及营销活动等。

（9）位置主数据

提供跟踪和共享地理信息的能力，并根据地理信息创建层次关系或地图。

- 1) 位置参考数据通常包括行政区域数据
- 2) 位置主数据包括业务方地址和位置

（10）行业主数据——参考目录

参考目录是主数据实体（公司、人员、产品等）的权威清单，组织可以购买和使用主数据实体作为交易的基础。虽然参考目录是由外部组织创建的，但管理并协调妥善的信息版本是在组织自己的系统中进行维护的。

参考目录可以通过以下方式帮助用户更好地使用主数据：

- 1) 为新纪录的匹配和连接提供起始点
- 2) 提供在记录创建时可能较难获得的其他数据元素
- 3) 当组织的记录与参考目录匹配、协调时，可信记录将偏离参考目录，并且可追溯到其他源记录，贡献属性和转换规则。

4、数据共享架构

1) 注册表

是指向多种记录系统中主数据记录的索引。记录系统管理应用程序本地的主数据，可以根据主索引访问主数据

2) 交易中心

各应用程序与中心系统交互，实现对主数据的访问和更新。

3) 混合模式

混合模式是注册表和交易中心的混合体。记录系统管理应用程序本地的主数据。主数据在一个公共存储库中被合并，并经由数据共享中心实现共享。



02

活动

1、主数据管理活动

1、识别驱动因素和需求

驱动因素包括改善客户服务和运营效率，以及减少与隐私和法律法规有关的风险

障碍包括系统之间在数据含义和结构上的差异。

2、评估和评价数据源

主数据管理的工作结果之一可能是通过评估现有数据的质量来改进元数据。

评估数据源的目标之一是根据组成元数据的属性来了解数据的完整性。

3、定义架构方法

主数据管理的架构方法取决于业务战略、现有数据源平台以及数据本身，特别是数据的血缘和波动性以及高延迟或低延迟的影响。

4、建模主数据

5、定义管理职责和维护过程

6、建立治理制度，推动主数据使用

2、参考数据管理活动

1、定义驱动因素和需求

驱动因素是运行效率和更高的数据质量

2、评估数据源

3、定义架构方法

4、建模参考数据

5、定义管理职责和维护流程

6、建立参考数据治理制度



03

工具和方法

数据整合工具、数据修复工具、操作型数据存储（ODS）、数据共享中心（DSH）、专门的主数据管理应用来实现。



04

实施指南

就绪评估/风险评估、组织和文化变革

1、遵循主数据架构

整合方法应考虑企业组织架构、记录系统的数量、数据治理实施、数据访问延迟的重要性以及消费系统和应用程序的数量。

2、监测数据流动（实现以下目的：

- 1) 显示数据如何在整个组织中共享和使用
- 2) 在管理系统和应用程序中识别数据的血缘关系
- 3) 辅助进行问题根本原因的分析
- 4) 展示数据整合和消费整合技术的有效性
- 5) 通过数据消费展示源系统的数据值延迟
- 6) 确定在集成组件中执行的业务规则和转换的有效性。

3、管理参考数据变更

数据修改的类型包括：

- 1) 对外部参考数据集的行级变更
- 2) 外部参考数据集的结构变化
- 3) 对内部参考数据集的行级变更
- 4) 内部参考数据集的结构变化
- 5) 创建新的参考数据集

4、数据共享协议

为确保恰当的访问和使用，应建立共享协议，规定哪些数据可以共享，以及在何种条件下可以共享。

5、组织和文化变革



05

参考数据和主数据治理

决定事项和度量指标

1、治理过程决定事项

- 1) 要整合的数据源
- 2) 要落实的数据质量规则
- 3) 遵守使用规则的条件
- 4) 要监控的行动和监控频率
- 5) 优先级和数据工作响应等级
- 6) 如何展示信息以满足利益相关方的需求
- 7) 参考数据管理和主数据管理部署的标准授权接口和预期。

1) 数据质量和遵从性

指标应该说明主题域实体或相关属性的置信度（百分比），以及它在整个组织中符合实际需求的使用价值。

2) 数据变更活动

指标展示数据值的变化率。

3) 数据获取和消费

指标应该显示和追踪哪些系统在贡献数据，哪些业务区域在共享环境中订阅数据。

4) 服务水平协议（SLA）

5) 数据管理专员覆盖率

指标应该关注对数据内容负责的个人或团队，并展示覆盖率的评估频率。

6) 拥有总成本

成本可以包括环境基础设施、软件许可证、支持人员、咨询费、培训等。
这一指标的有效性是基于其在整个组织中的持续应用。

7) 数据共享量和使用情况

指标展示数据共享环境中流入和流出数据的定义，纳入和订阅的数量和速率

考试资料职业发展
技术读书笔记分享

B站/闲鱼：大西洋活跃的锅巴
公众号：不太甜

本章完结 感谢观看

完整课程视频请扫描二维码咨询

