

数据标准培训

上海汉得信息技术股份有限公司
HAND Enterprises Solutions Co, Ltd.





1

数据标准的基本概念

2

数据标准的设计步骤

3

Q&A 环节

GB/T: 数据的命名、定义、结构和取值的规则。

-----（GB/T 36073-2018 数据管理能力成熟度模型）

华为: 数据标准（Business data standard）定义公司层面需共同遵守的属性层数据含义和业务规则，是公司层面对某个数据的共同理解，这些理解一旦确定下来，就应作为企业层面的标准在企业内被共同遵守。

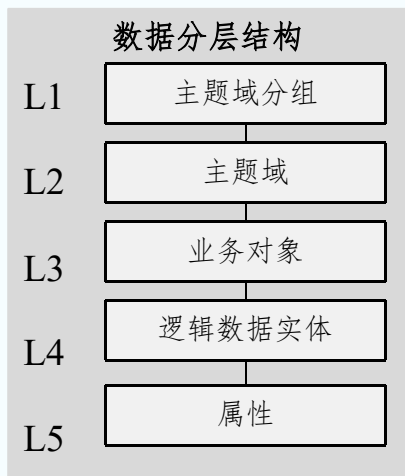
-----（《华为数据之道》）

信通院: 是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。

-----（《数据标准管理实践白皮书》）

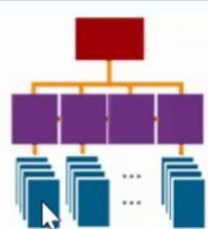
JR/T: 对数据的表达、格式及定义的一致约定，包含业务属性、技术属性和管理属性的统一定义。

-----（JR/T 0105-2014 银行数据标准定义规范）



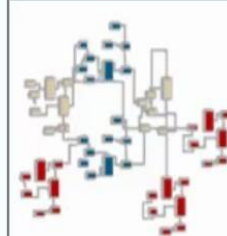
1 数据资产目录

- 通过分层架构表达对数据的分类和定义
- 厘清数据资产
- 是建立数据模型的输入

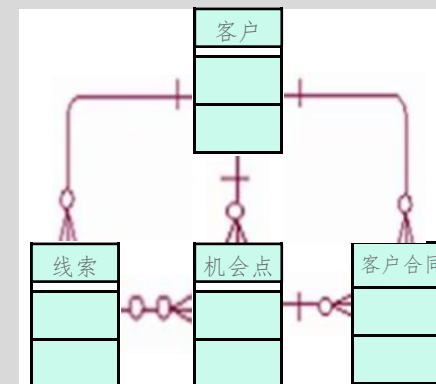


数据模型 2

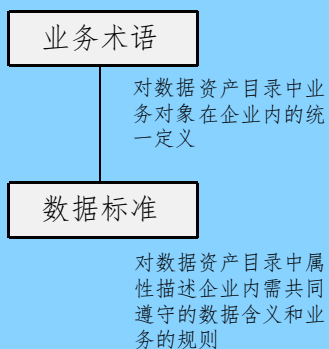
- 通过E-R建模实现对数据及其关系的描述
- 指导IT开发，是应用系统实现的基础



数据模型



数据标准组件



3 数据标准

- 是业务定义的规范
- 统一语言，消除歧义
- 为数据资产梳理提供标准的业务含义和规则



数据架构

数据分布 4

- 是数据在业务流程和IT系统上流动的全景视图
- 识别数据的“来龙去脉”
- 是定位数据问题的导航



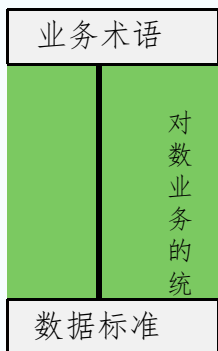
数据分布组件

表达数据在业务流的流转

表达数据在IT系统中的流转

定义数据产生的源头

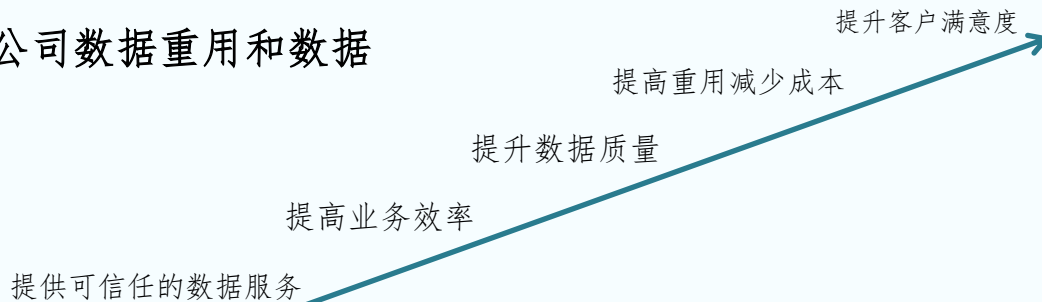
数据标准



据资产目录中
对象在企业内
一定义

对数据资产目录中
属性描述企业内需
共同遵守的数据含
义和业务的规则

*在公司层面统一数据定义，为公司数据重用和数据质量的提升奠定基础。



业务

1. 统一规则

- 业务术语和数据标准是业务运作和数据质量提升的基础；
- 数据标准中定义了标准格式、规则和允许值列表，为集成提供依据。

2. 统一定义

- 为数据项的含义在全公司提供了统一且一致的理解，业务术语和数据标准是公司标准，任何应用系统、流程文件都应该遵循。

3. 明确责任人

- 明确了数据标准的责任人，即数据管家，使得数据项的定义等问题可以找到责任主体。

4. 标准重用

- 在新开发过程中通过对标准的重用降低变革成本。

IT

身份证号的18位数分别代表：

- **1 1-2位**：省、自治区、直辖市代码；
- **2 3-4位**：地级市、盟、自治州代码；
- **3 5-6位**：县、县级市、区代码；
- **4 7-14位**：出生年月日，例如19670401代表1967年4月1日；
- **5 15-16位**：县、县级市、区代码；
- **6 17位**：性别标识，奇数代表男性，偶数代表女性；
- **7 18位**：作为尾号的校验码，由号码编制单位按统一的公式计算出来，可以是0-9的数字，也可以用X代替10。

员工专业占比分析



相关数据标准问题

- 专业名称不统一：
- 机电一体化、机电、机电一体化技术；
 - 计算机、计算机应用、计算机科学与技术；
 - 会计学、会计；
 - 电气自动化、电气工程及自动化；
- 以上本质为同一个专业，但是名称未统一。
- 专业名称易混淆：
- 电子（占比9%）容易和电子商务、电子信息工程、电子电工之间混淆，没有规范统一的名称。

带来的影响

- 降低工作效率：数据标准不统一可能导致在整合、处理和分析数据时需要额外的时间和资源。例如专业名称不统一导致在筛选简历范围时需要进行额外的判断、有可能遗漏目标专业的简历。
- 产生数据质量问题：不一致的数据标准可能导致数据质量问题。例如电子和电子商务、电子信息工程容易混淆，可能影响到数据分析的准确性和可靠性，从而对决策产生负面影响。
- 增加部门间沟通障碍：当不同组织或部门使用不同的数据标准时，沟通和信息交流可能会受到阻碍，导致误解、错误解读或无法理解其他方提供的信息。例如专业名称不规范，导致业务部门和HR的沟通成本增加，增加了招聘到错误专业人才的风险。

标准编码	标准名称	状态
<input type="checkbox"/>	标准编码	新建
<input type="checkbox"/>	SA03030026	在线
<input type="checkbox"/>	SA03030016	离线
<input type="checkbox"/>	SA03030014	发布审核中
<input type="checkbox"/>	SA03030010	下线审核中
<input type="checkbox"/>	SA03030009	采购组
<input type="checkbox"/>	SA03030008	采购请求中的批准策略
<input type="checkbox"/>	SA03030007	采购申请发布状态
<input type="checkbox"/>	SA03030006	采购申请批准标识
<input type="checkbox"/>	SA03030005	采购申请建立标识(采...
<input type="checkbox"/>	SA03030004	采购申请处理状态
<input type="checkbox"/>	SA03030003	采购凭证类别
<input type="checkbox"/>	SA03030031	采购申请凭证类型
<input type="checkbox"/>	SA03030030	采购申请总价值
<input type="checkbox"/>	SA03030029	采购申请申请数量
<input type="checkbox"/>	SA03030028	短缺数量(范围内库存)
<input type="checkbox"/>	SA03030027	采购申请技术规范名称
<input type="checkbox"/>	SA03030027	采购申请技术规范代码

用户诉求：将发布审核中、新建、在线、离线调整为编制中、审批中、已发布、已作废。

供应商答复：

- 1. 本次需求不属于之前确认的需求范围，将此需求作为优化项提交到产品端，等待产品端排期回复（是否能够作为优化项，暂时无法确认）。
- 2. 如需明确交付时间，视为定制开发，需要商务上讨论一下这块内容如何处理；
- 3. 此项内容虽然看上去不复杂，但是由于涉及到枚举值的修改，除了前端界面上的显示需要修改外，影响内容较多，工作量也会相对较高：
 - ① 标准的新建修改等保存逻辑
 - ② 审批时的交互逻辑
 - ③ 产品中的一些统计类的任务逻辑
 - ④ 导入导出逻辑需求变化
 - ⑤ 平台中目前涉及到状态的都需要做同样的处理，避免平台不统一。

- 数据标准包含业务、技术和管理三方面，数据标准设计分别从这三个方面进行设计。
- 参考数据包含参考数据信息、参考值信息两部分，在设计数据标准时同步梳理对应的参考数据。

数据标准					参考数据	
业务信息			技术信息	管控信息	信息	参考值信息
■ 用于统一业务语言和理解，明确定义每个属性所遵从的业务定义和用途、业务规则、同义词，并对名称进行统一定义，避免重复			■ 对IT实施形成必要的指引和约束，包括数据类型、长度、精度	■ 明确各业务部门在贯彻数据标准管理方面应承担的责任	■ 每个参考数据都应当明确定义其取值的数据标准来源	■ 每个参考值的取值、定义及用途
主题域分组	标准编码	数据示例	数据类型	责任部门	参考数据编码	顺序号
主题域	标准名称	同义词	数据长度	责任人	参考数据名称	参考值取值
业务对象	业务用途	标准依据	数据精度	责任人电话/邮箱	参考数据分组	参考值含义
逻辑数据实体	业务规则	依据内容	是否为空	标准适用范围	定义及用途	定义及用途
	引用参考数据编码			数据密级		父参考值取值

业务信息	*主题域分组	属性标准所属的主题域分组中文名称，与数据资产目录L1中文名称保持一致
	*Subject Domain Group	属性标准所属的主题域分组英文名称，与数据资产目录L1英文名称保持一致
	*主题域	属性标准所属的主题域中文名称，与数据资产目录L2中文名称保持一致
	*Subject Domain	属性标准所属的主题域英文名称，与数据资产目录L2英文名称保持一致
	*业务对象	属性标准所属的业务对象中文名称，与数据资产目录L3中文名称保持一致
	*Business Object	属性标准所属的业务对象英文名称，与数据资产目录L3英文名称保持一致
	*逻辑数据实体	属性标准所属的逻辑数据实体中文名称，与数据资产目录L4中文名称保持一致
	*Logical Data Entity	属性标准所属的逻辑数据实体英文名称，与数据资产目录L4英文名称保持一致
	*标准编码	属性标准的唯一识别编码，标准编码以SA(Standard Attribute)加10位数字构成，1-2位代表主题域分组编号，3-4位为主题域编号，后6位为流水号
	*标准名称	属性标准的中文名称，与数据标准唯一来源的数据资产目录L5中文名称保持一致
	*Data Standard Name	属性标准的英文名称，与数据标准唯一来源的数据资产目录L5英文名称保持一致
	*业务定义及用途	对属性标准的定义，解释这个属性标准是什么，对业务的作用，不能与中文名称相同或无业务含义的字符
	*Business Definition and Purpose	英文语言解释对属性标准的定义，解释这个属性标准是什么，对业务的作用，不能与英文名称相同或无业务含义的字符
	*业务规则	属性标准的业务规则，包括但不限于属性标准在各场景下的变化规则，编码含义等。详见业务规则填写示例
	*Business Rules	属性标准的业务规则英文版，包括但不限于属性标准在各场景下的变化规则，编码含义等。详见业务规则填写示例
	*引用参考数据编码	代码类数据标准需要填写所引用的参考数据对应的编码，如果没有允许值范围的限制，则填写“无需引用”。针对属性标准的取值有允许值范围的限定的情况，需要在参考数据模板中新建对应的参考数据，或者引用已发布的参考数据
	*数据示例	属性标准实例化的样例，用以帮助其他人员对此属性标准的理解
	同义词	业务对于同一属性可能有不同的称呼，在此列出业务对此属性的其它叫法
技术信息	*标准依据	制定数据标准项定义所依赖的参考依据，主要包括：国际标准、国家标准、行业标准、行业实践、企业标准、软件标准等，若无需引用标准，则填无参考依据
	依据内容	填写参考标准的具体的名称及编码，例如参考标准类型为国家标准，填写国家标准文件编码，参考标准为企业标准，填写企业发文的编码，除了无需引用标准的情况外，其他参考标准类型都需要填写参考标准
	*数据类型	业务定义的数据类型，与IT实现可能不一致，类型可选项：数值、文本、日期、代码、编码、时间、金额、比率，代码类的数据标准需要引用参考数据
	*数据长度	填写数字；根据实际业务需要，按照应录入的符号数量进行合理长度定义，此处所指符号包括但不限于中文汉字、英文字母、阿拉伯数字、标点符号等各类符号。根据如下规则对属性的数据长度限制进行合理定义： 对于属性录入内容长度可评 可预知，例如描述性或备注性的属性信息，则预留更大的长度，避免由于各类不确定业务因素导致需要频繁调整数据长度
管控信息	数据精度	填写数字，数值型、金额型、比率型数据标准项小数位所占字符个数或数字位数
	是否为空	数据是否可为空
	*数据密级	参考：《信息安全管理规范》规范 1、绝密：不正当揭露对公司财务或竞争力造成伤害；此类数据的泄漏或丢失至少会导致：失去重要的客户订单和采购价格优势、技术先机，造成严重的经济损失；违反国家法律法规，出现大量负面报道；对整个公司有严重的影响 2、机密：不正当揭露造成公司竞争上的困难或影响公司声誉；此类数据的泄漏或丢失至少会导致：损失少量客户部分少量订单或采购价格优势，造成合作伙伴或客户的投诉，造成一定的经济损失；违反合同相关约定或引起了媒体的关注；对一个或多个部门有一定的影响。 3、内部使用：不宜让不需要的人获知；对销售或采购、产品生产、日常管理造成较小的影响。 4、公开：可对内外部公开或法令规定须共公开。
	*标准适用范围	如果填集团，则要求全集团范围内遵从此数据标准。否则填写适用的BU/MU业务领域，代表该业务领域需要遵从此标准
	*责任部门	是指对业务规则的制定、解释、维护和执行负有最终责任、负责数据的日常维护、更新、校验、清洗和归档等工作的部门。责任部门需要确保业务规则的准确性、一致性和适用性，以便业务流程能够正确执行并且需要确保数据的准确性、完整性、一致性和时效性，以便数据能够支持业务决策和运营，可以按实际业务情况。填写制定数据标准的主责部门，建议写一个，如需多方确认，由责任主体部门牵头拉通。
	*责任人	是指责任部门中该数据标准具体的负责人。
	责任人电话/邮箱	是指数据标准负责人的联系方式，发现数据标准的定义或者业务用途描述有疑问等可快速联系到相应负责人