
聚力·赋能
2016阿里安全峰会
2016 ALIBABA SECURITY SUMMIT

数据治理与安全：从隔空相望到并肩前行

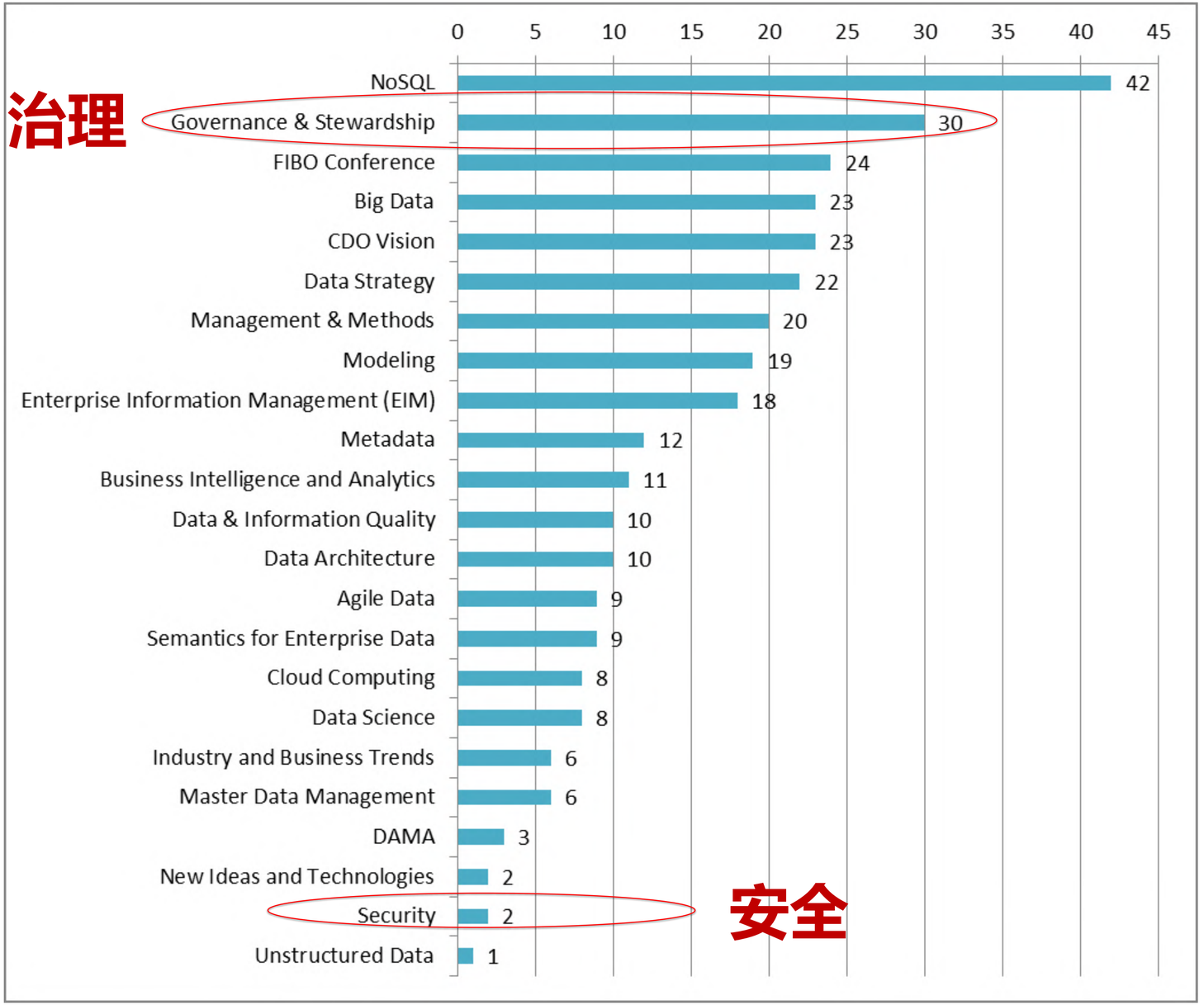
御数坊 刘晨

个人简介

刘晨，数据治理与管理领域专家、御数坊创始人。

- 国际数据管理协会中国分会（DAMA China）副会长，清华大数据产业联合会副秘书长；
- 拥有多年通信、金融、能源等行业数据治理规划与实施经验；
- ITSS数据治理标准工作组成员、信标委大数据标准化工作组成员；
- 获得EDME、CDMP、DGP、IQCP等四项国际认证；
- 出版物：《DAMA数据管理知识体系指南》、《首席数据官实战》等。

为什么选择这个主题？



2016年4月，美国San Diego，EDW企业数据世界论坛，为期6天，200+议题

- 企业数据战略
- 数据治理项目实施
- 构建数据架构的新需求
- 数据质量度量和计分卡
- 大数据趋势和技术
- 主数据管理
- 企业信息管理 - 数据驱动的业务变革
- 实时分析和商业智能
- 企业数据管理全方位最佳实践
- 非关系型数据库

为什么选择这个主题？

数据治理与数据安全的隔空相望

- 数据治理：数据模型，数据标准，数据质量，元数据，主数据，参考数据，**数据安全**...
- 数据安全：网络安全，系统安全，数据库安全，**数据安全**...

然而，从知识到人员，数据治理与数据安全却总似尤抱琵琶半遮面...

目的：加强交流，并肩前行！

议题

- 概念解读：数据管理、数据治理与数据安全
- 实践回顾：数据治理的国内外案例分享
- 行动思考：基于数据标准化的数据安全管理思路

为什么需要数据治理？

不良数据治理导致的损失...

- **元数据不一致导致卫星失败：**
  - NASA，1999年，火星气候探测器，任务失败。
  - 因为火星气候探测者号上的飞行系统软件使用**公制单位牛顿**计算推进器动力，而地面人员输入的方向校正量和推进器参数则使用**英制单位磅力**，导致探测器进入大气层的高度有误，最终瓦解碎裂。

- **编码不一致产生的ERP建设返工、工期拖延：**
  - 国内某大型企业实施SAP，SAP ECC模块公司代码与BCS模块贸易伙伴并行维护导致数据不一致，评估**返工工作量>1000人天**

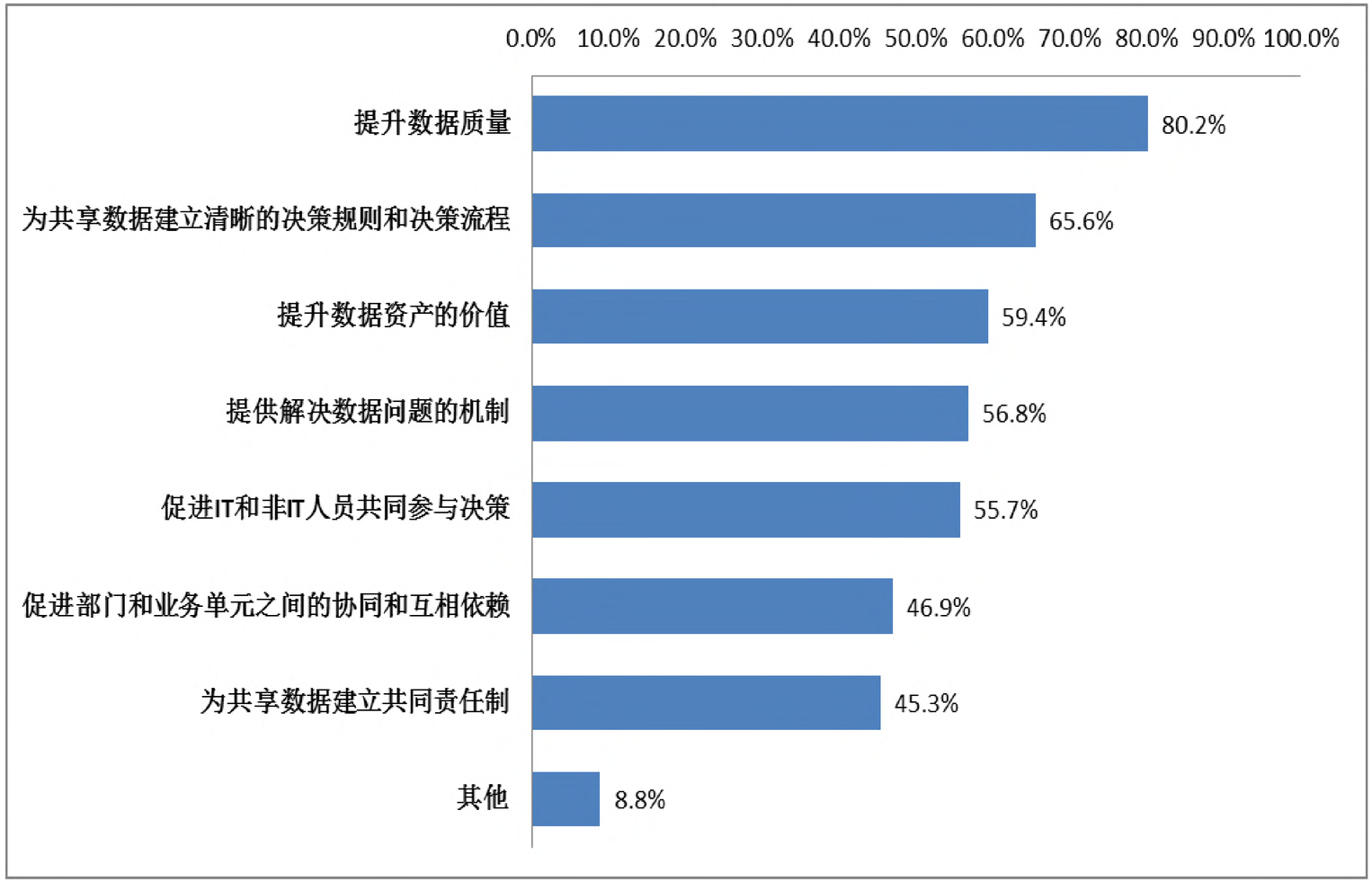
- **不良数据质量导致的财务损失：**
  - 根据数据质量专家Larry English的统计，截至2010年，不良数据质量为122家知名机构带来的财务损失总计：

$1, 212,374,479,000

数据引自《Information Quality Applied》

开展数据治理的背景和价值

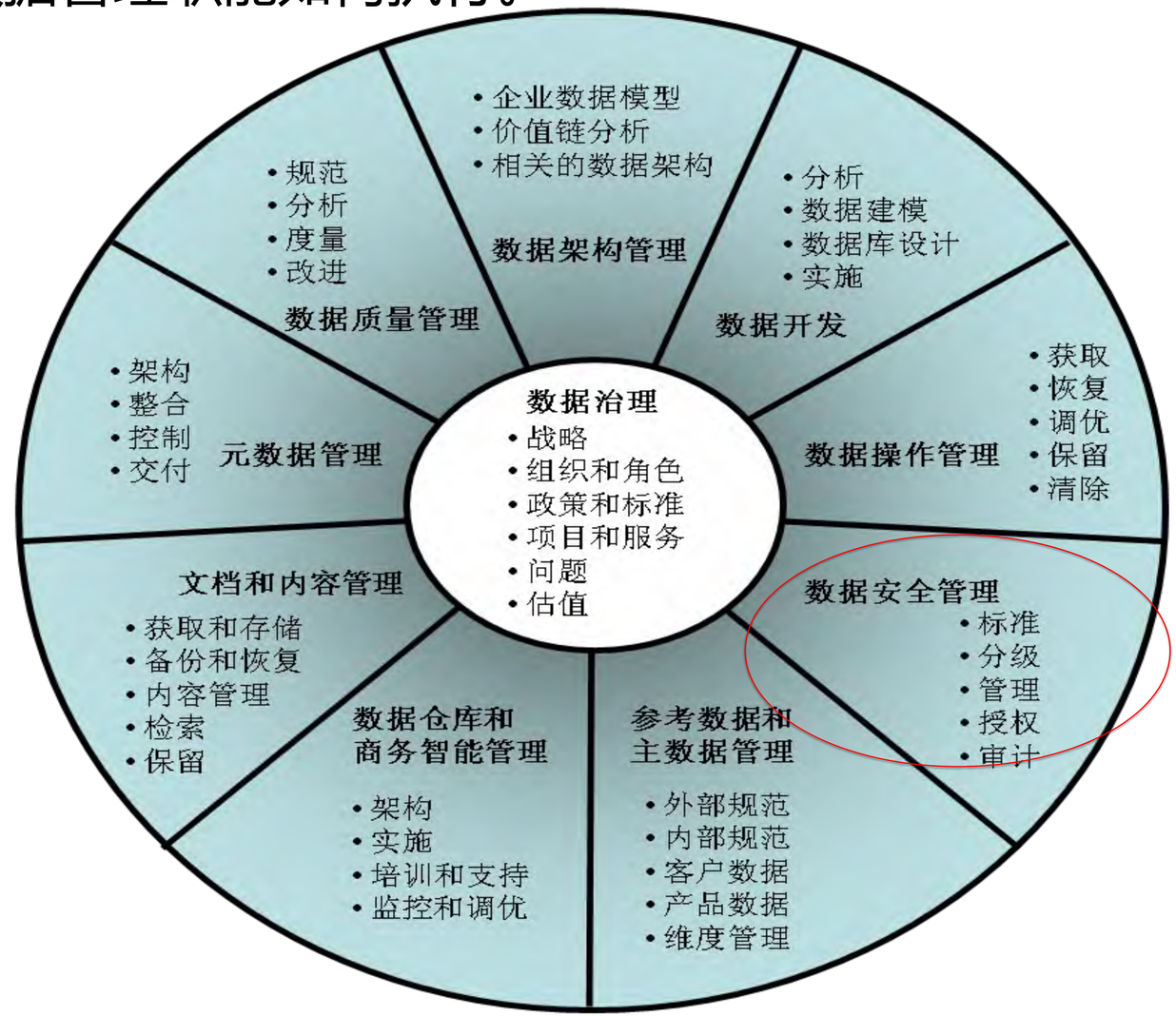
- **背景**
  - 合规问题
  - **安全问题**
  - 兼并收购
  - 大数据
  - 企业数据仓库/BI建设
  - 数据集成/SOA
  - 数据质量问题和影响
  - 主数据管理
  - 公共业务术语



注：调研数据来自国际数据质量协会

什么是数据治理？

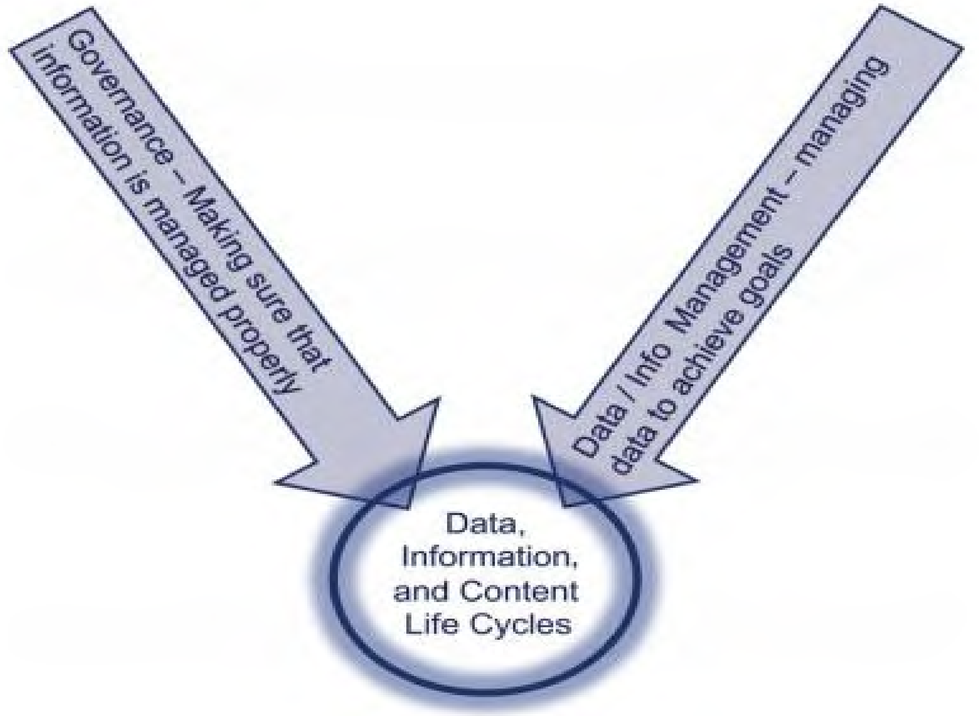
- 数据治理是**对数据资产的管理**行使权力和控制的活动集合（规划、监控和执行）。数据治理职能指导其他数据管理职能如何执行。



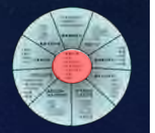
数据治理与数据管理的区别：广义与狭义

- **数据管理（DM）：**
  - **规划、控制和提供**数据和信息资产，发挥数据和信息资产的价值。
  - **EIM：强调企业级；DM可以在企业级和局部进行**
- **数据治理（DG）：**
  - **对数据资产管理活动（the management of data assets）行使权力和控制的活动集合**（规划、监控和执行）。数据治理职能指导其他数据管理职能如何执行。
- **数据治理制定正确的原则、政策、流程、操作规程，确保以正确的方式对数据和信息进行管理。**

Governance= management of management



数据治理，谁之责？



领导：数据不对？从哪里出的？

业务部门：从系统出的！

IT部门：...

领导：IT部门负责查明原因，解决问题！

IT部门：...

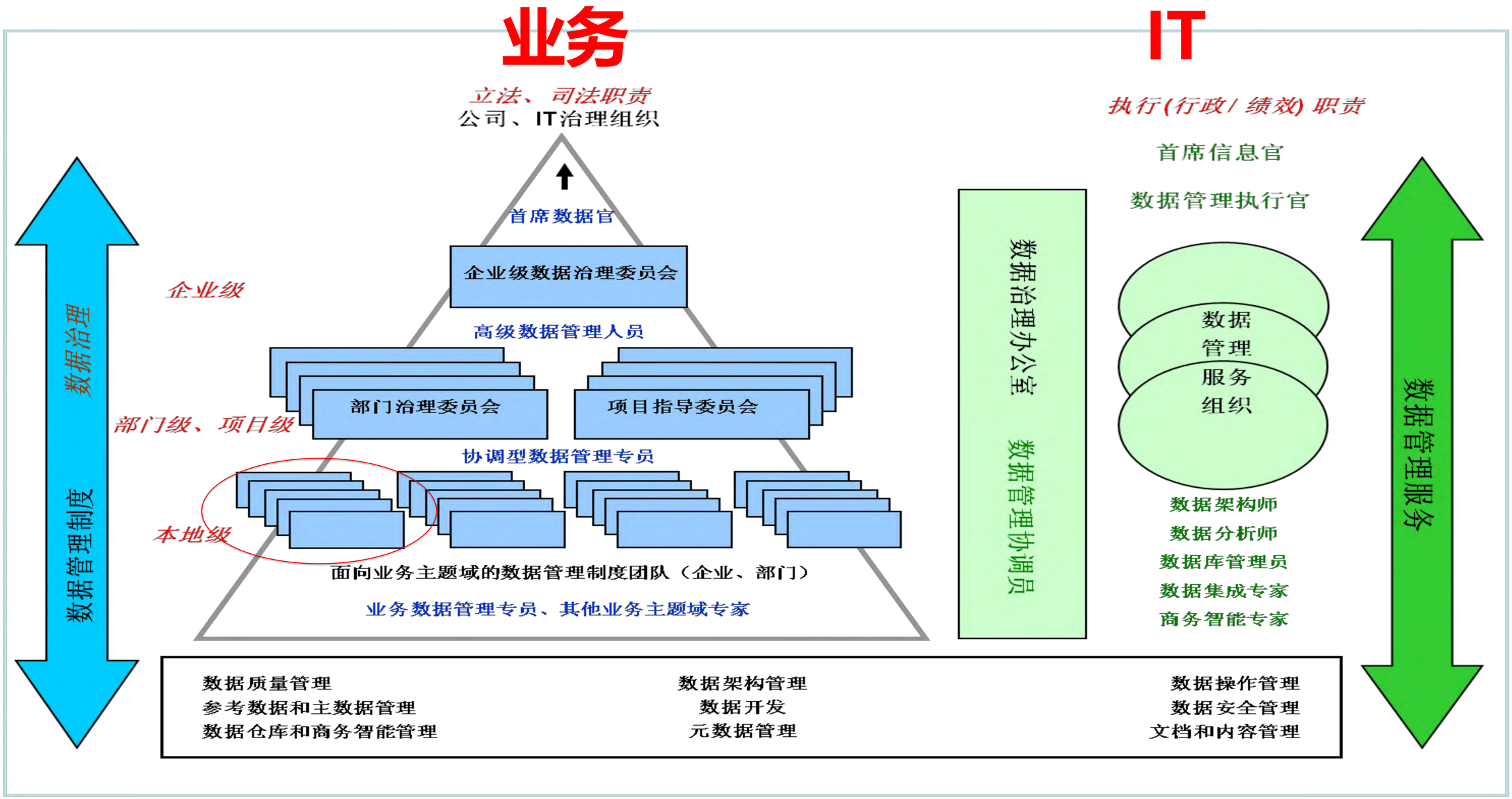
于是，数据的问题，被广泛的认为是IT部门的职责。

而IT部门，也饱受其苦：

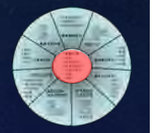
- *数据定义和业务规则，业务部门最清楚；*
- *数据录入，业务人员负责；*
- *数据使用，业务人员是用户；*
- *数据考核，业务部门有权力*

数据治理，是业务部门与IT部门共同的职责

数据治理的组织架构：强调不同领域业务人员的深度参与和决策



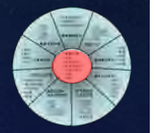
数据管理专员制度：Data Stewardship



- 数据管理专员 Data Steward
  - –Steward：管家，翻译成管家不够严肃，因此采用了“专员”。
  - –Steward与Owner相对应，说的是**虽然资产不是归Steward所有，但是他们替Owner代管**

- 数据管理专员制度 Data Stewardship
  - –也衍生出Stewardship一词，表明代管、托管制度
  - –数据管理专员制度主要**探讨业务部门应承担的数据管理角色、职责以及相应的能力要求和制度。**

什么是数据管理专员制度？
Data Stewardship



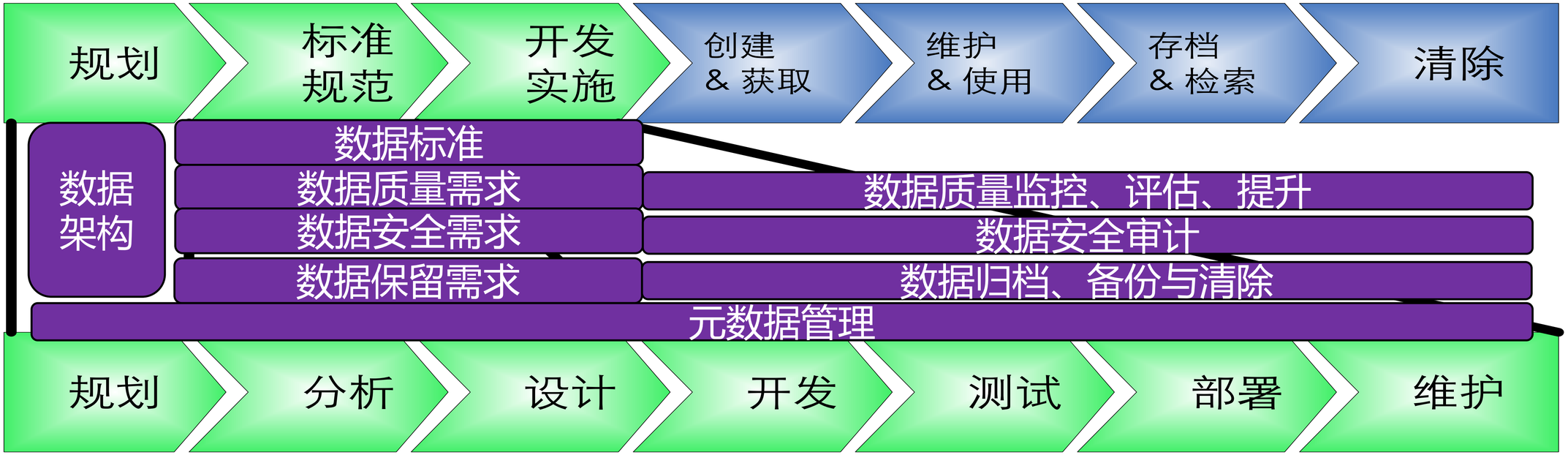
- **数据管理专员制度是为数据资产管理分配的、委托的业务职责和正式的认责**
  - –是数据管理工作在业务方面的职责，对应的还有IT数据专业人员的职责。

- **数据管理专员（Data Stewards）定义和监视数据的定义、质量、访问和保留**
  - –数据治理——对“如何管理数据”进行决策
  - –定义业务数据的名称、业务含义
  - –定义和维护参考数据值
  - –定义业务数据需求
  - –识别和解决数据问题
  - –定义数据质量需求和度量指标
  - –监测数据质量
  - –定义主数据管理和数据衍生计算的业务规则
  - **–定义某些数据安全和访问规则**
  - –定义某些数据保留规则和规程

数据治理方法：面向数据生命周期
时间视角—数据在IT系统建设和应用中被影响

- 企业管理数据资产，就是管理数据的生命周期。
- 数据先被创建或获得，然后存储、维护和使用，最终被销毁。
- 有效的数据管理，数据的生命周期开始于数据获取之前，企业先期制定数据规划、定义数据规范，以期获得实现数据采集、交付、存储和控制所需的技术能力。

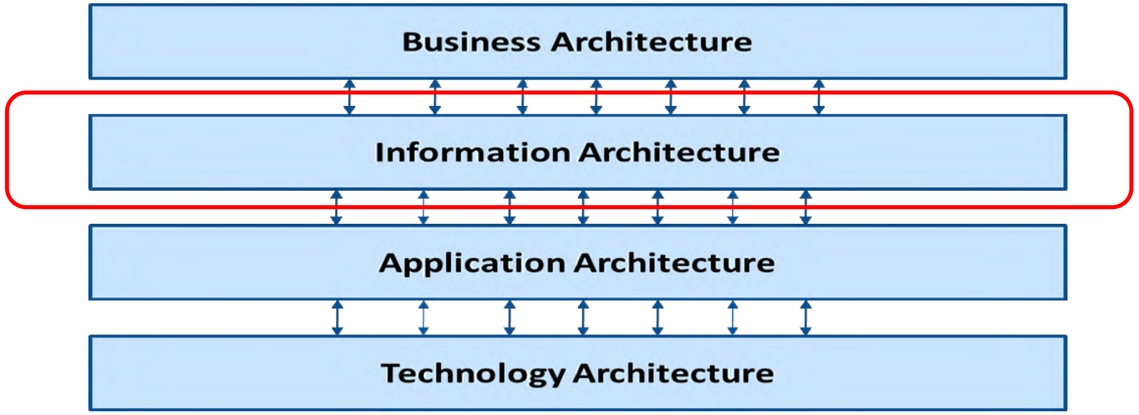
数据生命周期



系统开发生命周期 (SDLC)

© DAMA China, 2015

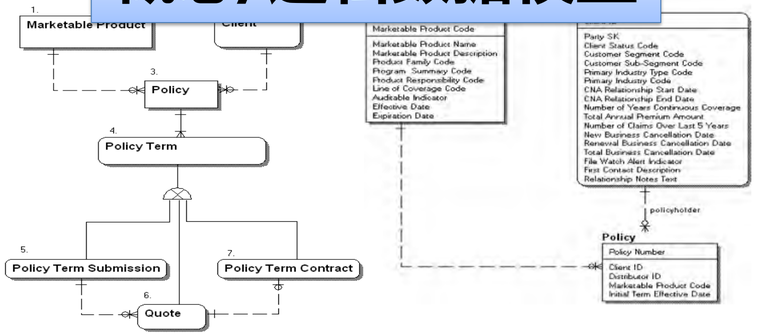
数据架构：以企业数据模型为中心



- 企业数据架构一般包括三套主要设计组件：
  1. **企业数据模型，**识别主题域、业务实体、控制实体元素之间关系的业务规则，以及若干重要的业务数据属性。
  2. **信息价值链分析，**使数据模型组件（主题域和/或业务实体）与业务流程及其它企业架构组件相一致；这些组件可能包括组织、角色、应用、目标、战略、项目和技术平台。
  3. **相关的数据交付架构，**包括数据技术架构，数据整合架构、数据仓库/商务智能架构、企业对内容管理的分类方法，以及元数据架构。

- 企业数据架构不仅仅涉及数据，它还采用通用的业务术语来帮助建立企业内的语义。

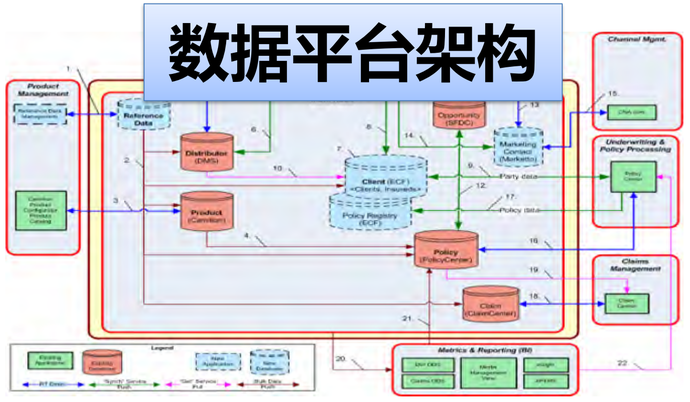
概念/逻辑数据模型



信息价值链分析

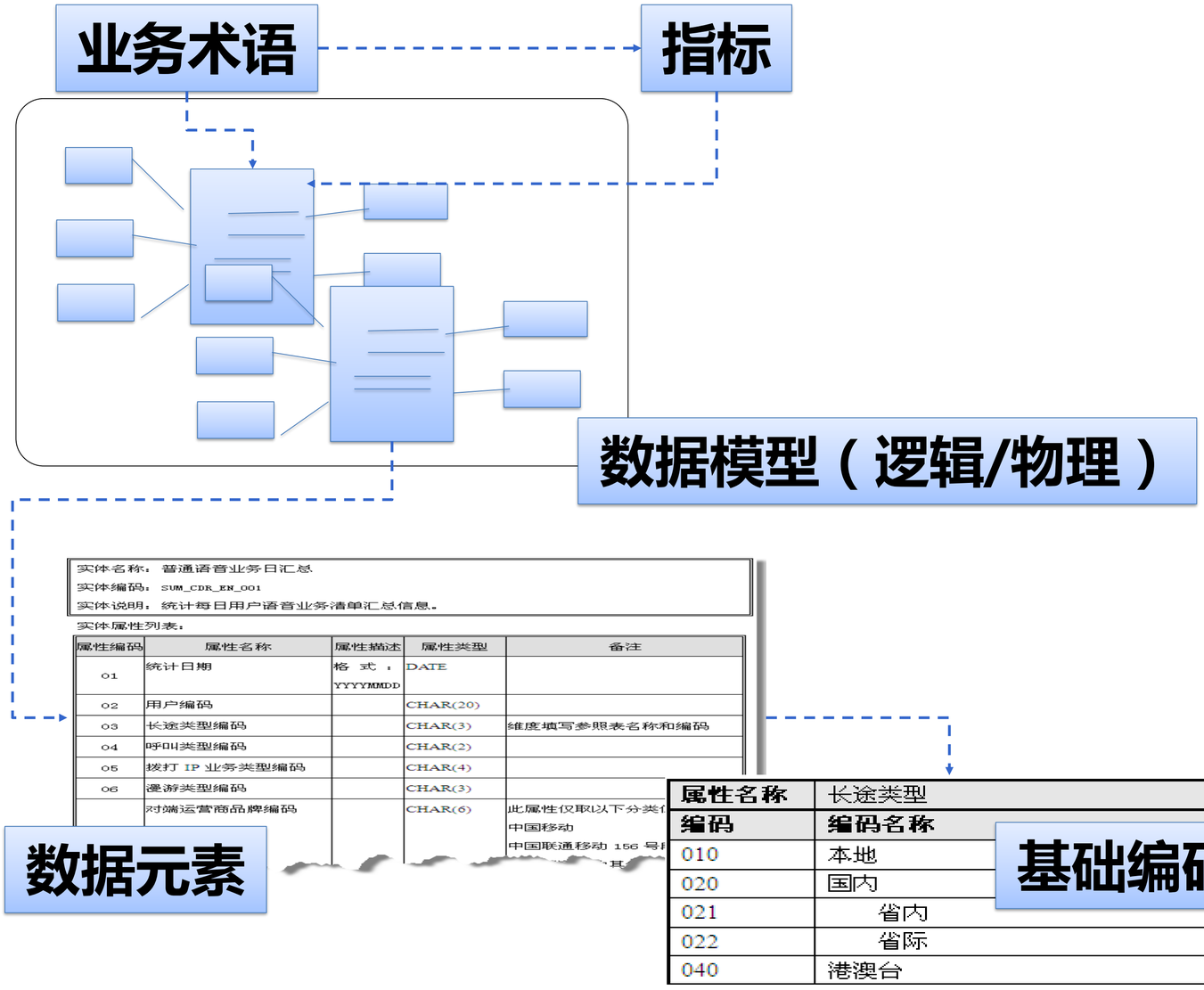
Business Capabilities / Subject Areas (C=Create, R=Read, U=Update, D=Delete)	Staff and Orga	Products	Producers (Ag	Clients (Custon	Policies	Exposure	Premium	Reinsurance	Claims	Legal Matters	Financial Assets	Technologies
Develop Products	R	C	R	R	R	R	R	R	R	R	R	
Market Products	R	R	C	C	R	R	R	R	R	R		
Manage Customer Relationships	R	R	U	U	R	R	R	R	R	R		
Underwrite Risks	R	R	R	U	C	C	C	R	R	R		
Collect Premium	R	R	R	R	R	R	U					
Process Claims	R	R	R	R	R	R	R	U	C	C		
Manage Risk Exposure	R	R	R	R	R	U	R	C	R	R		
Manage Financial Resources	R	R	R	R	R	R	R	R	R	R	C	R
Manage Human Resources	C	R	R	R	R	R	R	R	R	R	R	R
Manage Information and Technology	R										U	C
Provide Corporate Services	R	R	R	R	R	R	R	R	R	R	U	R

数据平台架构



从数据模型出发，延伸到不同粒度的数据标准

- **业务术语标准化：**
  - 业务概念、业务含义
  - 举例：客户、用户
- **指标标准化：**
  - 名称、含义、口径、计算逻辑、来源…
  - 举例：营业收入 vs 销售收入
- **数据模型标准化：**
  - 核心业务对象：员工，客户，产品，地址…
  - 举例：员工（员工号，部门号，姓名，性别，电话…）
- **数据元素标准化：**
  - 数据库字段：最小的数据粒度
  - 举例：身份证号码（中文名称，英文名称，数据类型，长度，有效取值…）
- **基础编码标准化：**
  - 对某个数据元素的取值定义
  - 举例：订单状态：已提交，已处理，退回，关闭…



定义数据质量需求

- **数据质量需求：**
  - –数据质量需求通常隐含在**业务政策**之中，**描述数据是否符合“适用性”（Fitness for Purpose）需求。**

- **数据质量维度包括：**
  - –准确性（Accuracy）
  - –完整性（Completeness）
  - –一致性（Consistency）
  - –时效性（Currency）
  - –精确度（Precision）
  - –隐私（Privacy）
  - –合理性（Reasonableness）
  - –参照完整性（Referential Integrity）
  - –及时性（Timeliness）
  - –唯一性（Uniqueness）
  - –有效性（Validity）

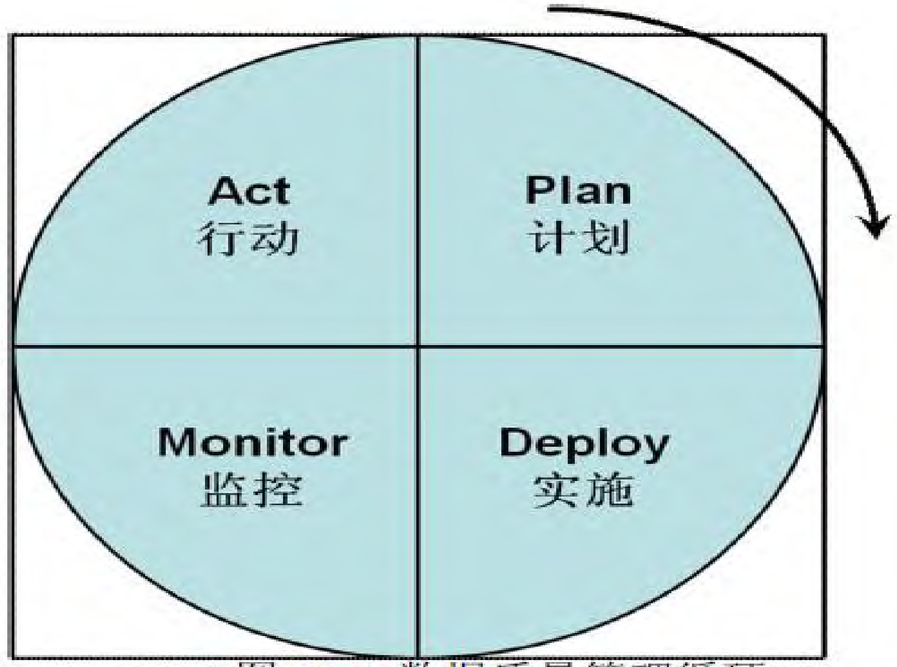


图 12.2 数据质量管理循环

数据安全管理的主要活动

5. 数据安全管理

定义: 通过计划、发展并执行数据安全政策和措施，为数据和信息提供适当的认证、授权、访问和审计。

目标:
1. 为数据资产读取和变更提供适合的方法、阻止不适合的方法。
2. 实现监管对隐私性和机密性的要求。
3. 确保实现所有利益相关者隐私性和机密性需求。

输入:
- 业务目标
- 业务战略
- 业务规则
- 业务流程
- 数据战略
- 数据隐私问题
- 相关的IT政策和标准

活动:
1. **理解数据安全需求和监管要求 (P)**
2. **定义数据安全策略(P)**
3. **定义数据安全标准 (P)**
4. **定义数据安全控制和措施 (D)**
5. **管理用户、密码和用户组成员 (C)**
6. **管理数据访问视图和权限 (C)**
7. **监控用户身份认证和访问行为 (C)**
8. **划分信息密级 (C)**
9. **审计数据安全 (C)**

主要可交付成果:
- 数据安全政策
- 数据机密和保密标准
- 用户档案、密码和成员资格
- 数据安全权限
- 数据安全控制
- 数据访问试图
- 文档分类
- 许可和访问历史
- 数据安全审计

供给者:
- 数据管理专员
- IT指导委员会
- 数据管理制度委员会
- 政府机构
- 客户

参与者:
- 数据管理专员
- 数据安全管理员
- 数据库管理员
- 商务智能分析师
- 数据架构师
- 数据管理领导
- 首席信息官/首席技术官
- 帮助台分析师

工具:
- 数据库管理系统
- 商务智能工具
- 应用框架
- 身份管理技术
- 变更控制系统

消费者:
- 数据生产者
- 知识工作者
- 经理
- 中高级管理人员
- 客户
- 数据管理专业人员

活动: (P) – 计划 (C) – 控制 (D) – 开发 (O) – 操作

- **理解数据安全需求和监管需求**
- **定义数据安全策略、标准、控制和措施**
- **管理用户、密码、用户组成员**
- **管理数据访问视图和权限**
- **监控用户身份认证和访问行为**
- **划分信息密级**
- **审计数据安全**

数据安全管理关键活动示例

- **理解数据安全需求**
  - –业务规则和流程定义了安全接触点。业务工作流中的每一个事件都有自身的安全要求。
    - **数据到流程（Data-to-Process）和数据到角色（Data-to-Role）**关系矩阵，引导数据安全角色、参数和权限的定义。
  - –需要在每个系统开发项目的分析阶段就识别具体的应用安全要求

- **关系数据库视图的数据安全机制**
  - –基于数据值将数据表中的数据限制到某些行。
  - –视图还可以允许广泛地获取某些列，并对密级更高的列限制访问。

数据安全管理关键活动示例：划分信息密级

- **划分信息密级：依据DAMA DMBOK**
  - **公众级**：信息可以提供给任何人，包括普通公众。一般公众级为默认分类。
  - **内部使用**：信息限制在雇员或组织成员中，但如果共享到其他人，风险也不大。通常仅供内部使用，可演示或讨论，但组织以外不得复制。
  - **机密**：资料不应共享至组织外部。客户机密信息不应该共享给其他客户。
  - **受限机密**：信息受限，承担某些角色的个人按需知密。
  - **注册机密**：这类信息如此机密以至于任何接触该信息的人都必须签署一份法律协议才能访问数据，并承担保密责任。

- **分级实施：**
  - 为数据库、表、列和视图分级。**信息密级的划分是元数据的最重要属性，指导如何赋予用户存取权限。数据管理专员负责数据密级的评估和确定工作。**

数据管理成熟度评估模型DMM

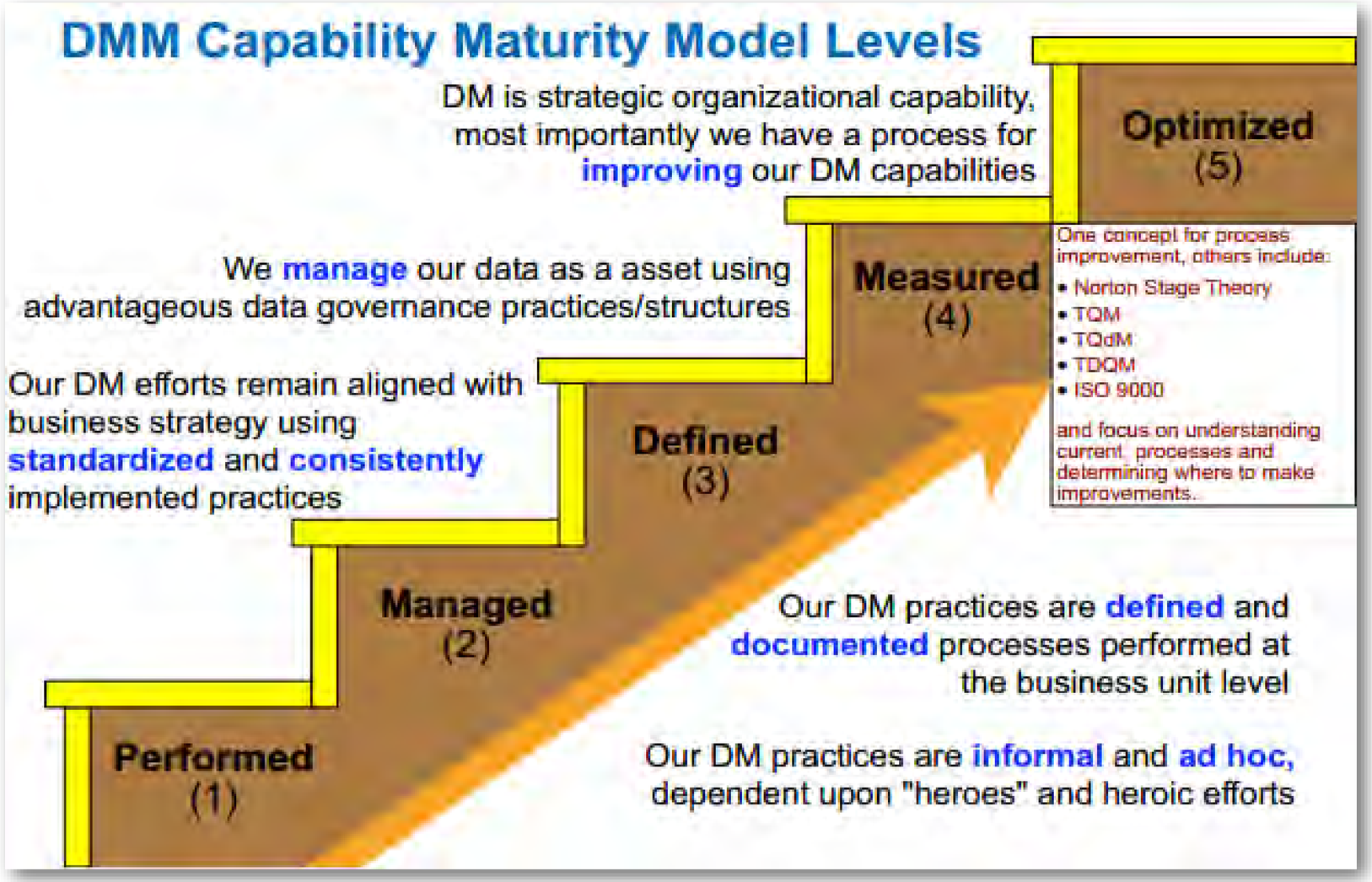
由于大数据技术的不断发展、成熟以及数据重要性的不断提升，美国的SEI（卡耐基梅隆软件工程研究所，是国际软件行业最通行的标准：CMMI的研发和推广方）在基于CMMI的方法和经验之上，结合众多知名厂商在数据管理领域的经验，于2014年8月正式推出数据管理成熟度评估模型（DMM）。

- DMM发布于2014年8月7日
- 3.5年开发时间
- 4个赞助机构：
  - Microsoft
  - Lockheed Martin
  - Booz Allen Hamilton
  - Kingland Systems
- 50+编写者
- 70+评审者
- 300+实践要点
- 500+工作成果

5大过程域25个子过程域

DATA MANAGEMENT STRATEGY	Data Management Strategy
	Communications
	Data Management Function
	Business Case
	Program Funding
DATA GOVERNANCE	Governance Management
	Business Glossary
	Metadata Management
DATA QUALITY	Data Quality Strategy
	Data Profiling
	Data Quality Assessment
	Data Cleansing
DATA OPERATIONS	Data Requirements Definition
	Data Lifecycle Management
	Provider Management
PLATFORM & ARCHITECTURE	Architectural Approach
	Architectural Standards
	Data Management Platform
	Data Integration
	Historical Data, Archiving and Retention
SUPPORTING PROCESSES	Measurement and Analysis
	Process Management
	Process Quality Assurance
	Risk Management
	Configuration Management

成熟度等级遵循CMMI划分，通过一系列核心提问识别企业数据管理实践水平



- 举例：数据管理战略的核心问题：

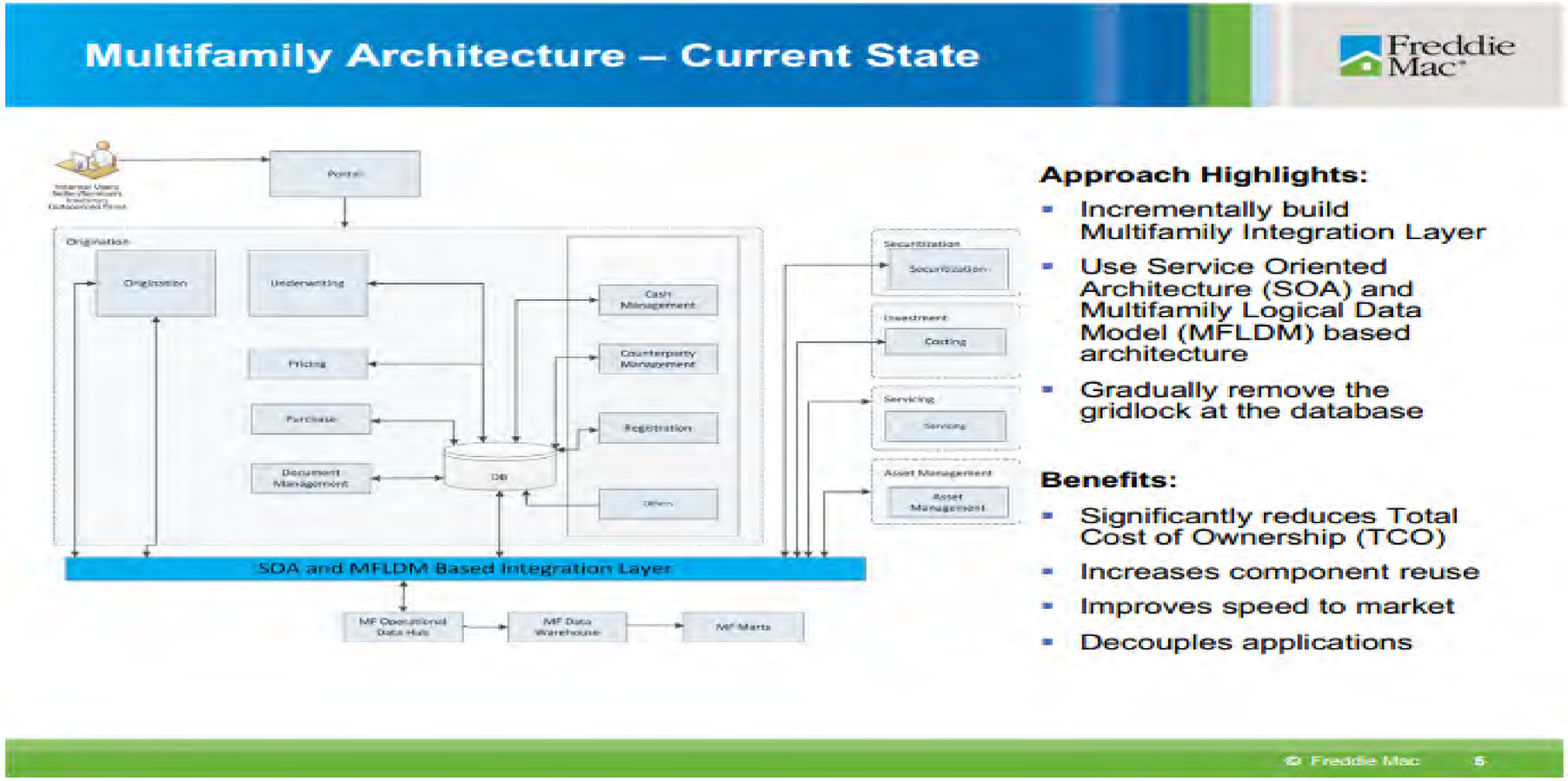
- 数据管理战略是否得到了高管层的显性和主动支持？
- 建设路线图是否与业务优先级和里程碑保持一致？
- 为支持长期可持续的数据管理计划，高管层、运营层、IT、业务相关方是否达成了充分理解和共识？
- 项目如何与建设演进路线保持一致？
- 员工能力和资源是否已就位？
- 是否有培训支持？

议题

- 概念解读：数据管理、数据治理与数据安全
- 实践回顾：数据治理的国内外案例分享
- 行动思考：基于数据标准化的数据安全管理思路

案例1：房地美（Freddie Mac）基于企业级数据模型实现数据整合共享

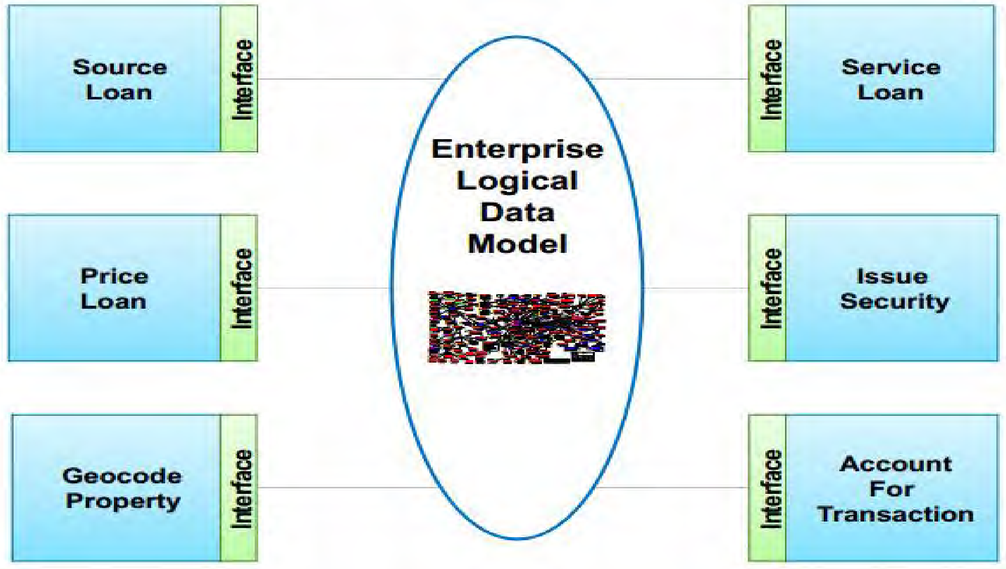
使用SOA架构+企业逻辑数据模型支撑企业数据集成，显著降低TCO，提高组件复用度，加快产品和服务对市场的响应，实现应用解耦。



房地美（Freddie Mac）案例：来源于业务需求、业务数据字典，在关键系统落地

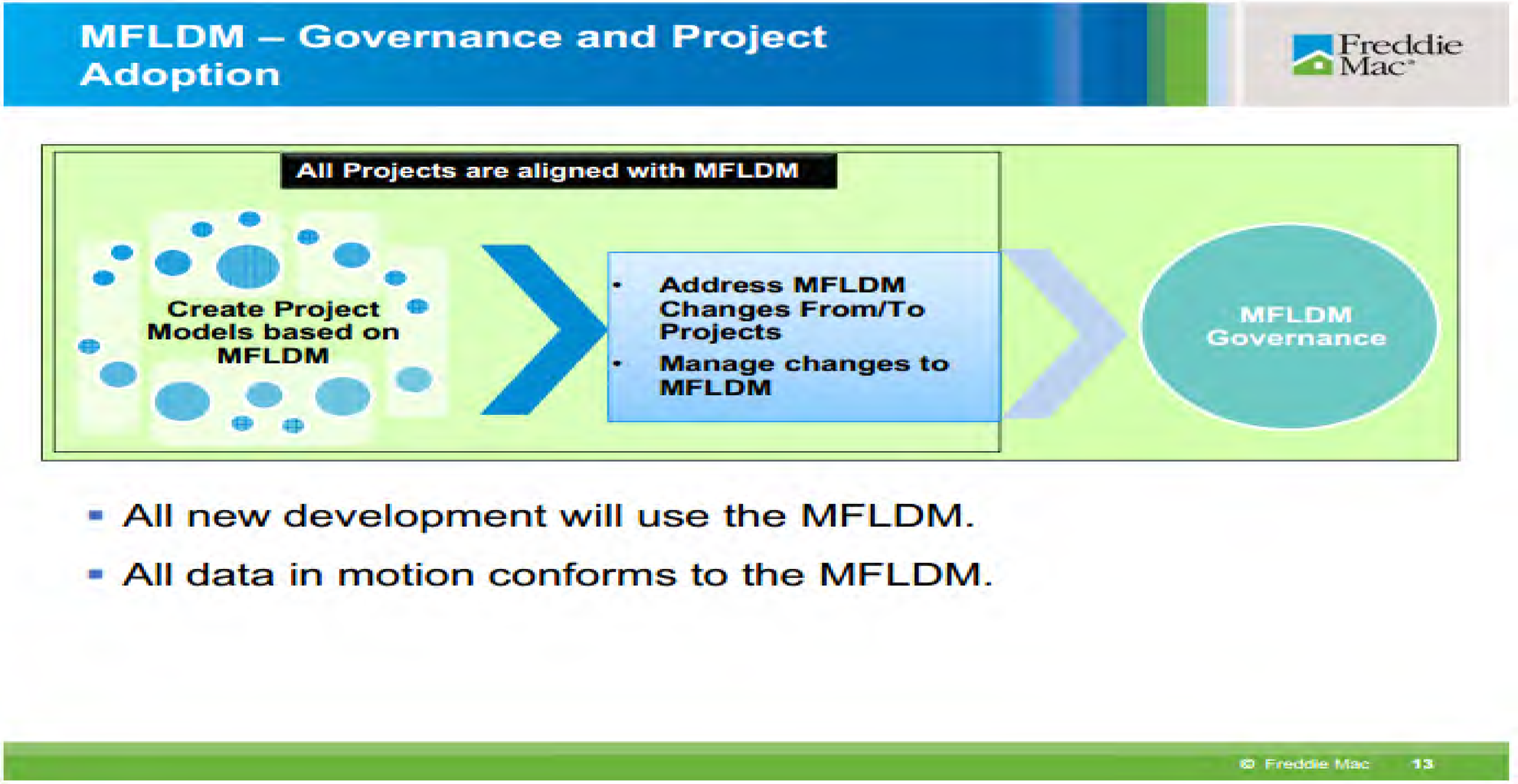
- The MFLDM:
  - Is an Enterprise Logical Data Model based on the Multifamily Business Data Dictionary (BDD)
  - Incorporates data requirements from in-flight business projects
  - Has a strong business sponsorship and is a joint effort between IT and the Multifamily organization
  - Has had three major releases
  - Is implemented in several strategic systems (including ODS and Data Warehouse)

- **基于业务数据字典**
- **考虑业务项目的数据需求**
- **高层支持、IT和业务部门合作**
- **3个主要版本**
- **在多个战略性系统中实施（ODS，DW）**



房地美（Freddie Mac）案例：在增量应用建设中管控，保障规范性

在项目建设过程中加强对数据架构和模型的管控，所有新开发项目必须使用企业数据模型，所有数据集成和移动必须遵从企业数据模型。



案例2：国内运营商，系统级数据治理

水源地治理

河道监管

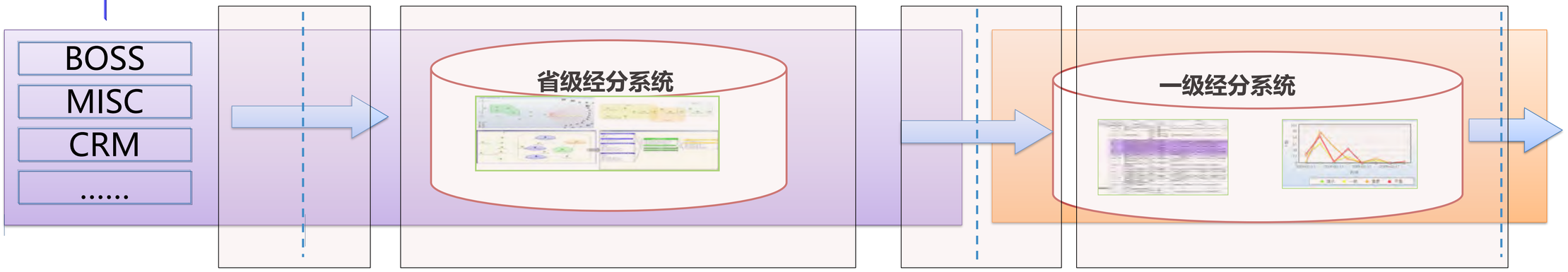
水厂监控

省经分与数据源系统的协同管理

数据质量管理工作类似 “水污染治理”

省级经分数据上报过程监控

一经数据质量全面监控

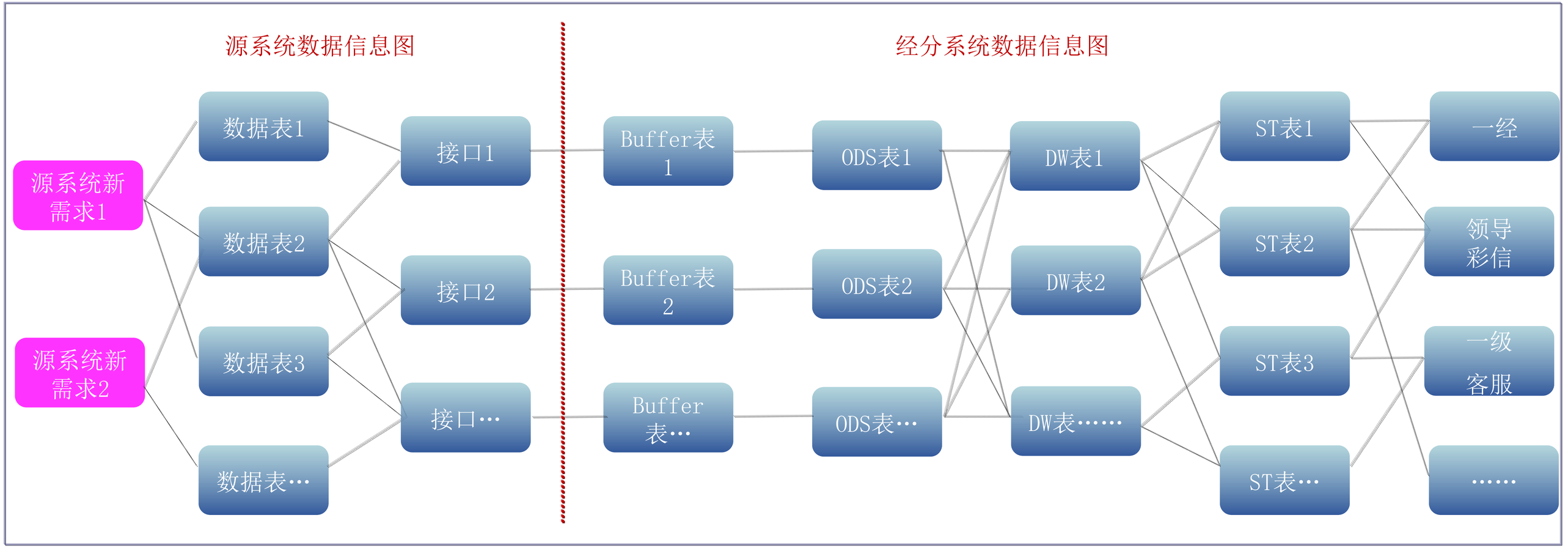


总部高层领导

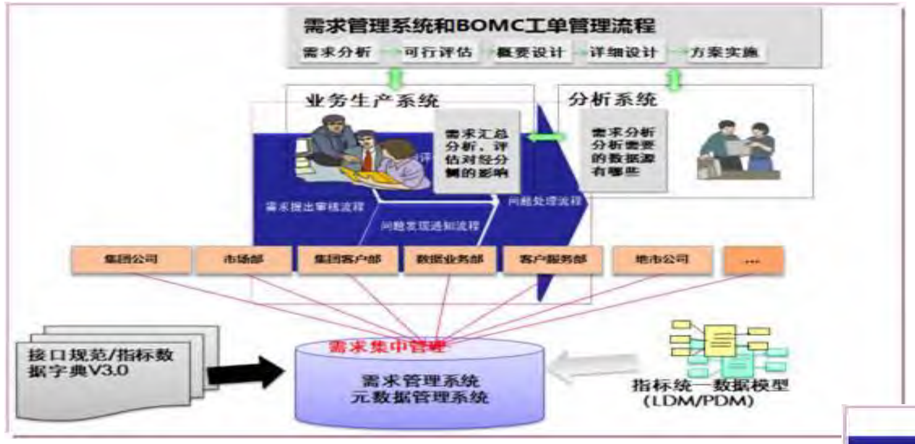
总部业务部门

数据源系统协同管理：构建支撑系统元数据信息地图

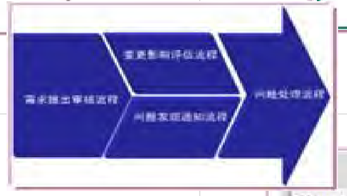
建立经分系统与源系统全局元数据信息地图---建立源系统新部署的业务需求、源系统数据库表、BI数据接口及经分应用之间的元数据信息地图，最终形成业务支撑网的全局信息地图。

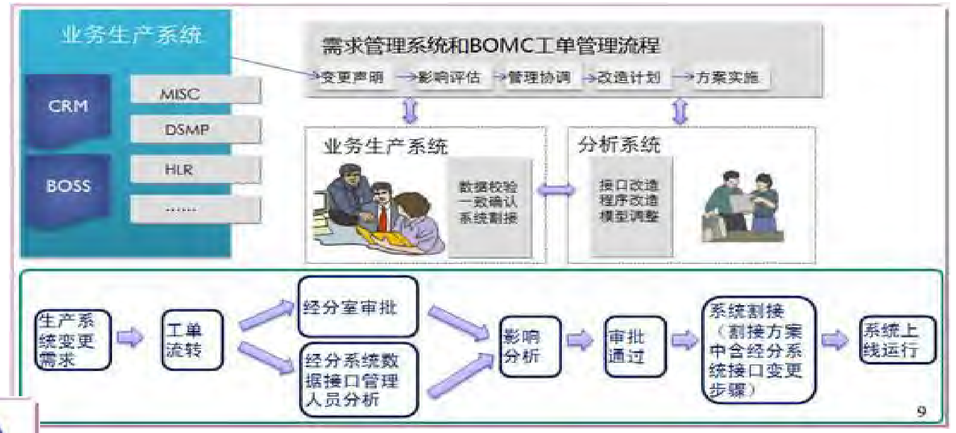


数据源系统协同管理：建立协同管理保障流程

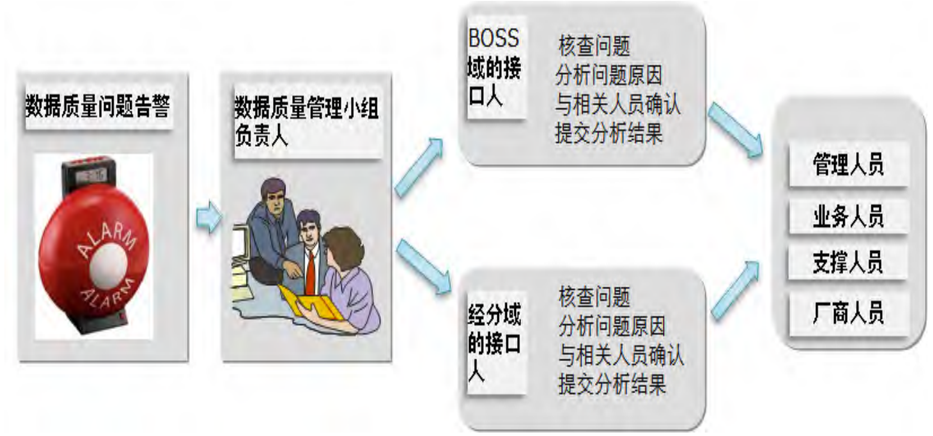


需求提出审核流程

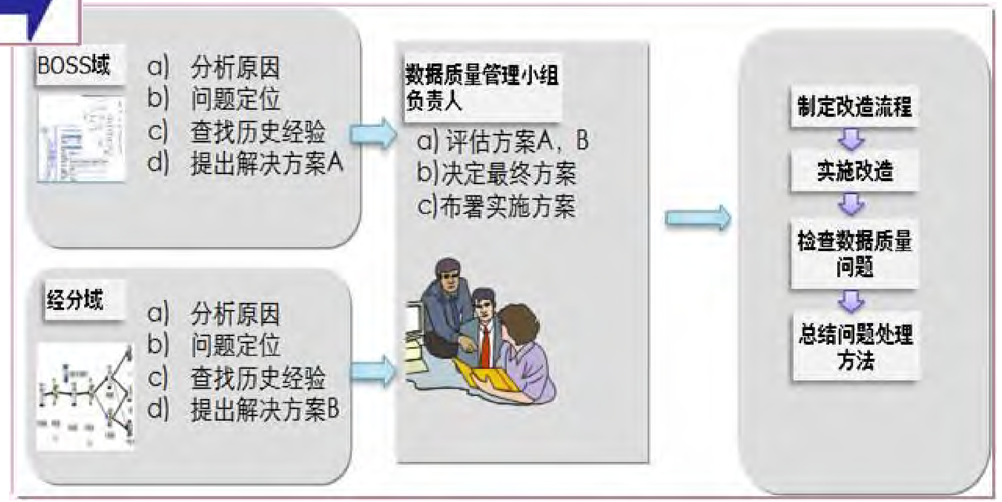




变更影响评估流程

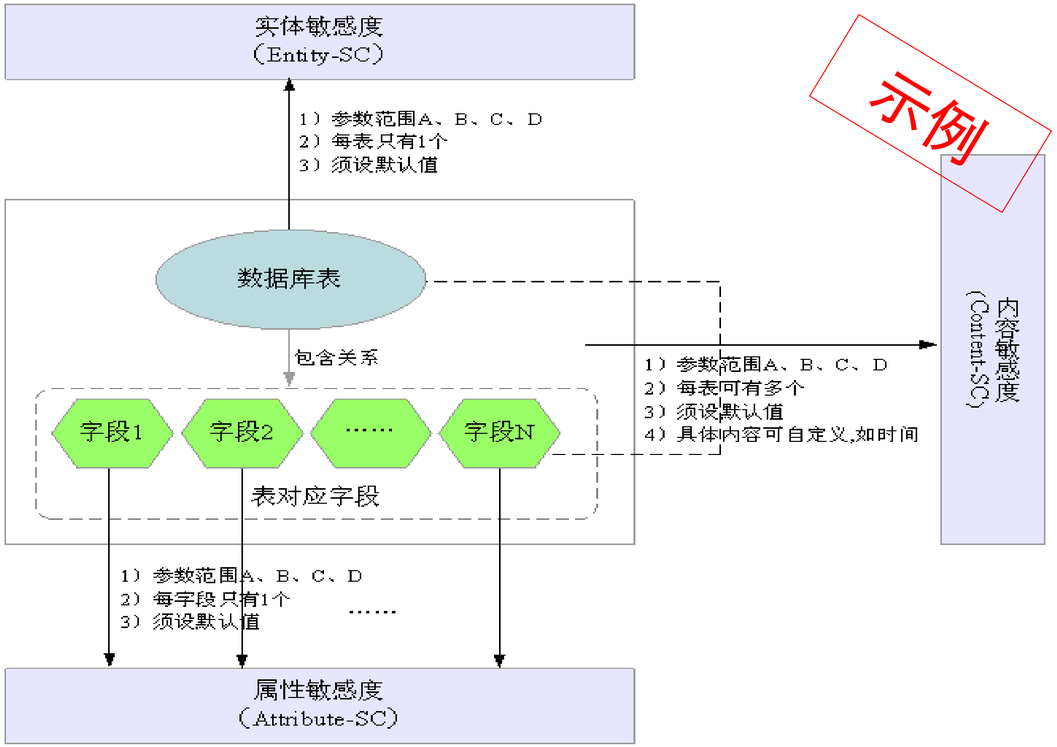


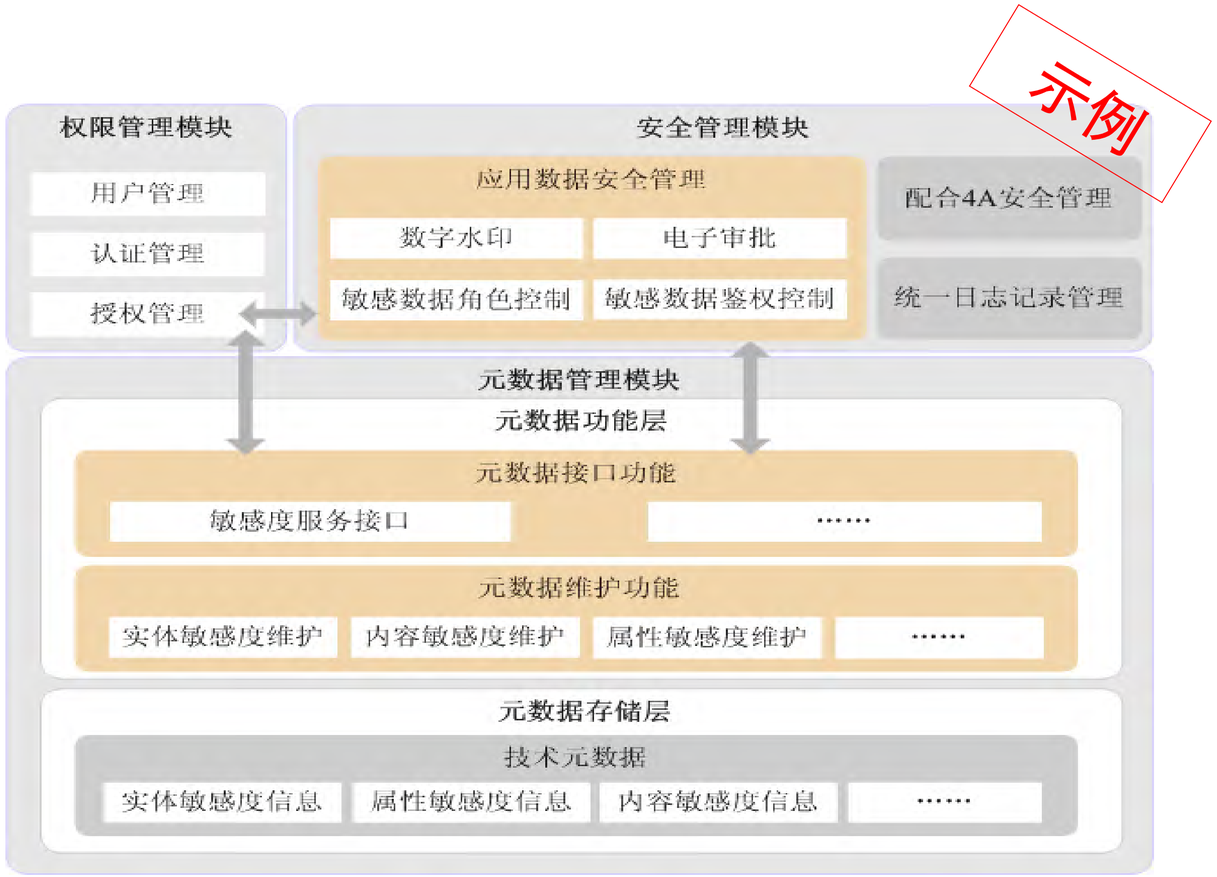
问题通知流程



问题处理流程

基于元数据实现数据安全敏感度管理





实践中的问题：

- **业务参与不足，IT不了解数据安全政策**
- **重功能，轻机制**

议题

- 概念解读：数据管理、数据治理与数据安全
- 实践回顾：数据治理的国内外案例分享
- 行动思考：基于数据标准化的数据安全管理思路

隔空相望...

已经做到的：

- 明确数据认责
- 梳理数据资源
- 梳理数据流程
- 制定数据标准

没做到的：

- 业务部门参与不足
- 不理解信息安全策略、未制定数据安全策略
- 未能将细粒度的数据资源、数据标准有效的与数据安全管理要求、技术手段关联

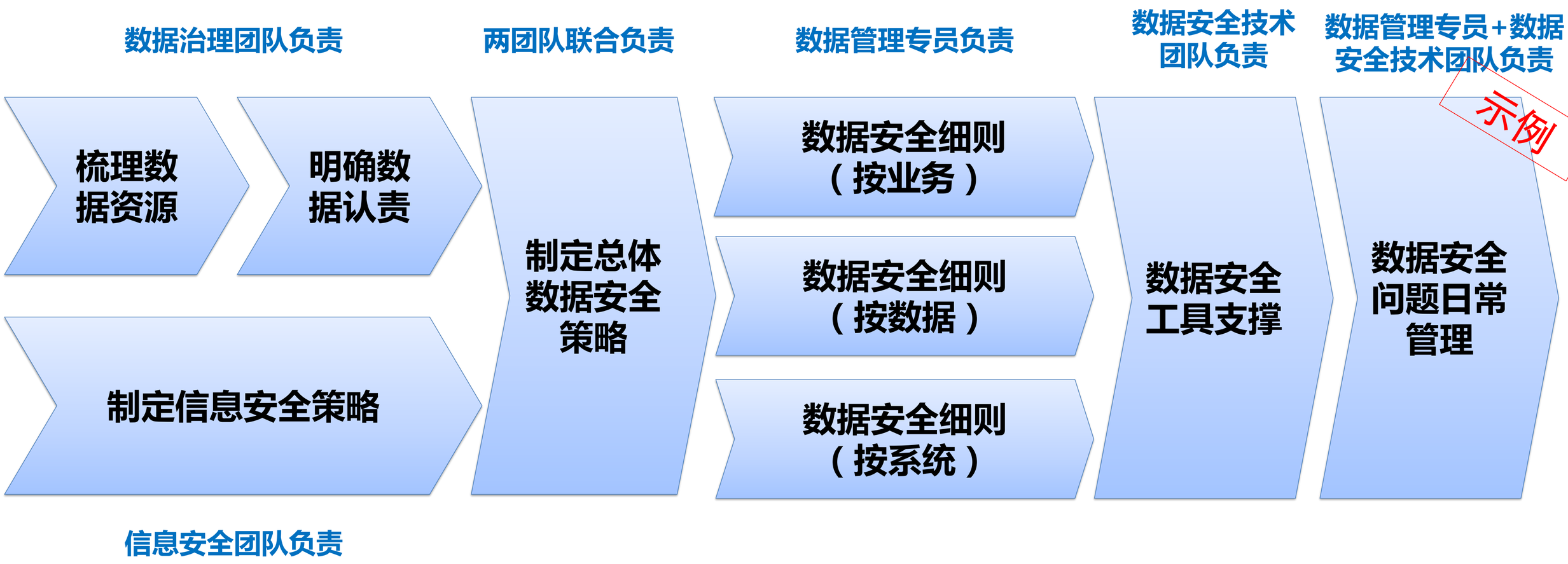
已经做到的：

- 相信有很多，传统数据治理人还需要学习...

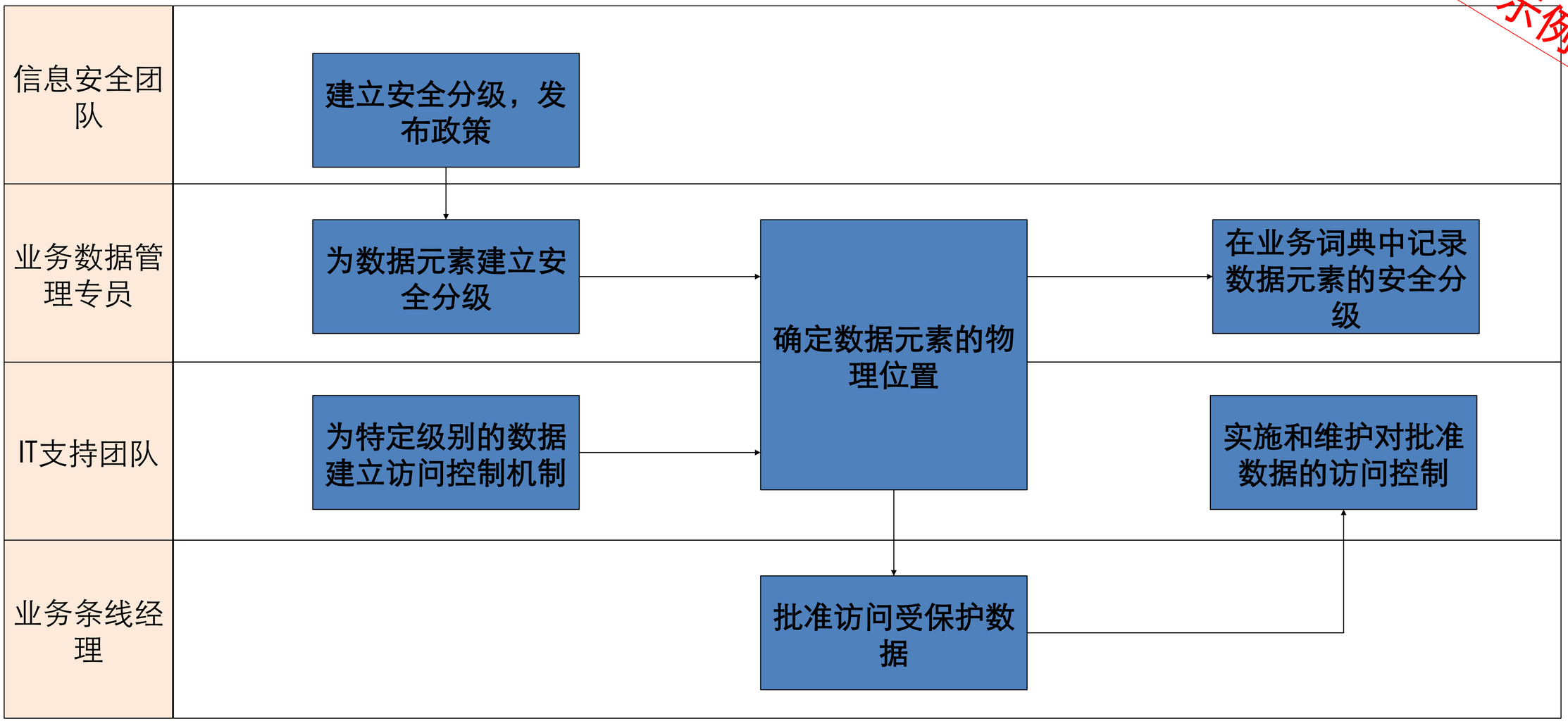
没做到的：

- 我们可以探讨数据治理能否、如何支撑数据安全？

如何并肩前行？


数据治理团队负责
两团队联合负责
数据管理专员负责
数据安全技术团队负责
数据管理专员+数据安全技术团队负责
梳理数据资源
明确数据认责
制定信息安全策略
制定总体数据安全策略
数据安全细则（按业务）
数据安全细则（按数据）
数据安全细则（按系统）
数据安全工具支撑
数据安全问题日常管理
示例
信息安全团队负责

示例：数据安全分级的职责和流程



示例：数据分类分级结果

示例

项目	数据主题	参考条目	参考数据级别	参考系统级别	备注
ERP 财务	总账管理、应收、应付管理、资金管理、预算管理、合并报表等数据主题	财务及资产管理规（计）划/财务报表	普通商密级	普通商密级	
ERP 物资	物资主数据、采购、库存、分析数据等数据主题	无	公开级	公开级	
ERP 人力	人事管理、薪酬管理、考勤管理等数据主题	干部统计表、信息库、花名册、收入、分配方案劳动工资统计年报等	普通商密级	普通商密级	

分享，互助，共赢

御数坊
Data Governance Workshop

御数有道，独具匠心
感谢聆听，敬请指正