



IBM 2010 新锐洞察高峰论坛

心·睿 掌控制胜先机





元数据如何帮助你的业务

James Wang
Sales Specialist, IPS
Software Group



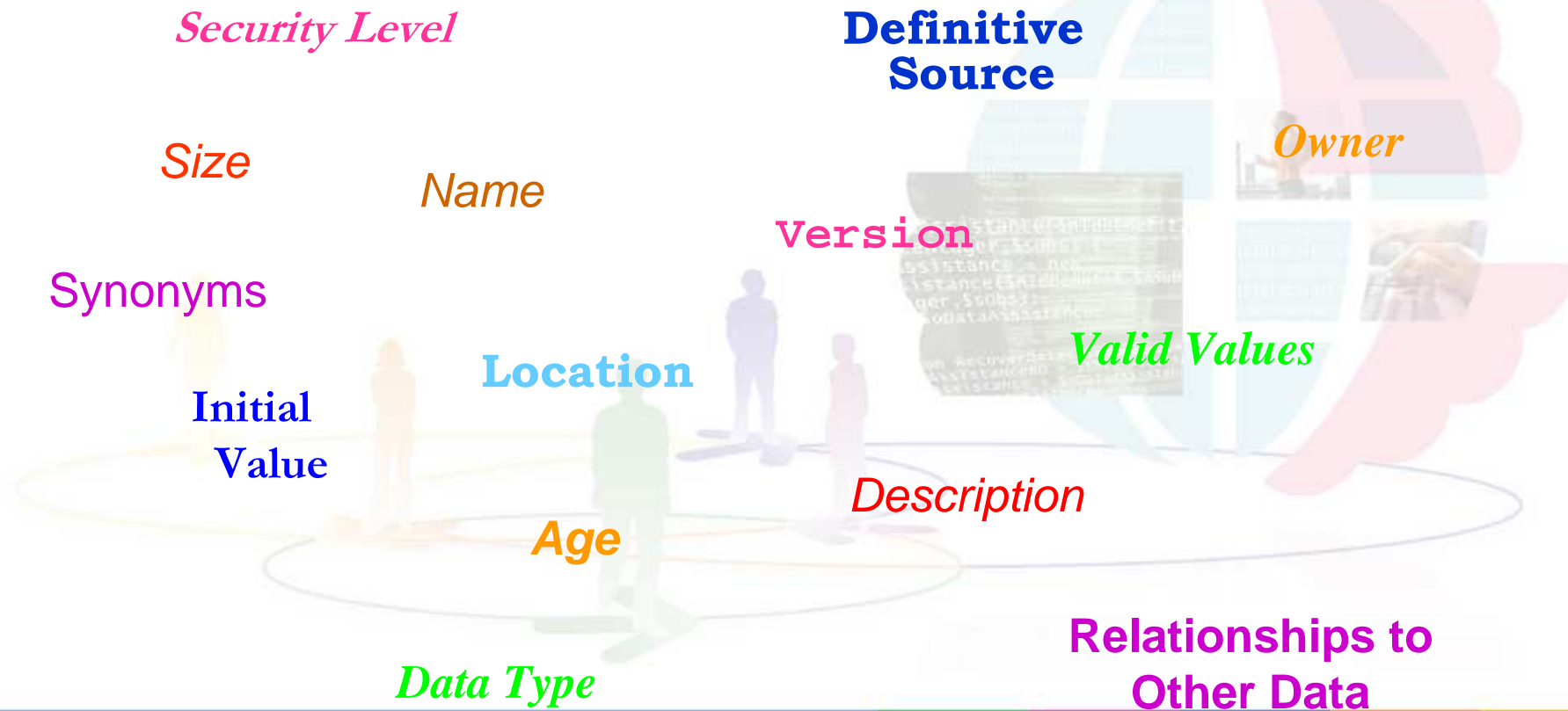


什么是元数据？



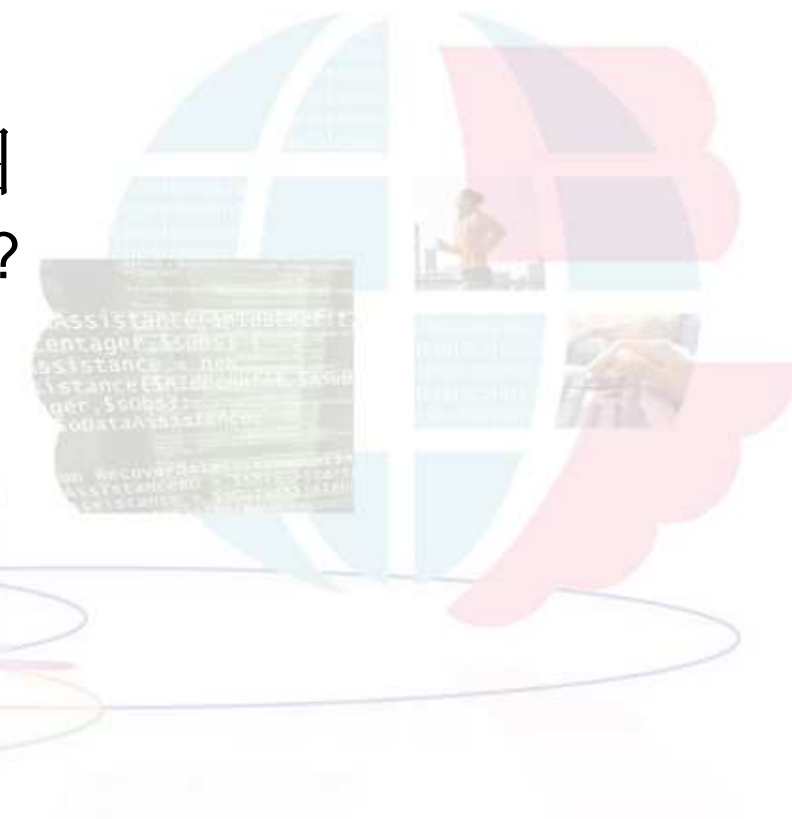
什么是元数据?

定义: Metadata - data about data
关于数据的数据



什么是元数据？ 另一个观点....

- 增加的语境(来龙去脉)
- 为了知识的导出提供路线图
 - 该数据从哪里来,要到哪里去?
 - 在哪里,如何我能找到它?
 - 它对我的意义是什么?
- 有关信息资产的数据



元数据到处都有

与任何事物与任何流程都有关系



元数据问题的风险和相关事物

- “巨大的交流问题造成该问题已被解决的假象。”*
- 当假象被粉碎后, 感到害怕, 不确定, 怀疑和受挫折.....
- 很明显你的数据是不一致的, 自相矛盾的或错误的从而导致:
 - 过多的开销:
 - 因为你不能找到你所需要的数据
 - 由于错误的决定产生了错误的数据
- 无法估计因改变而造成的花费和时间

元数据的类型

B

- **业务元数据**

- 业务规则、定义、术语、术语表、运算法则和系统使用业务语言。观众：业务用户

T

- **技术元数据**

- 定义源和目标系统以及它们的表和字段结构、属性、出处和依赖性。观众：特殊工具用户 – BI, ETL, Profiling, Modeling.

O

- **操作元数据**

- 应用程序运行信息：他们的频率、记录数以及各个组件的分析和统计信息。观众：操作管理和业务用户。

P

- **项目元数据**

- 文件和审计开发成就、代理干事和操作变化管理。观众：操作、干事和工具用户和管理。

字面意思, “**data about data**” 是 **从业务和技术的观点** 描述企业信息。





业务元数据

- 提供业务含义给**IT**的资产
- 赋予业务的含义给那些由其他的**IT**应用创建和管理的产物
- 较好的利用**IT**来实现业务的目标
- 被用来
 - 定义一个共享的数据的含义
 - 建立具有责任性的,可被审计的和可被跟踪的机制
 - 可控制的访问权限
 - 在用户中共享知识和经验
- 必须由哪些能理解信息资产对业务的意义和重要性的人来管理



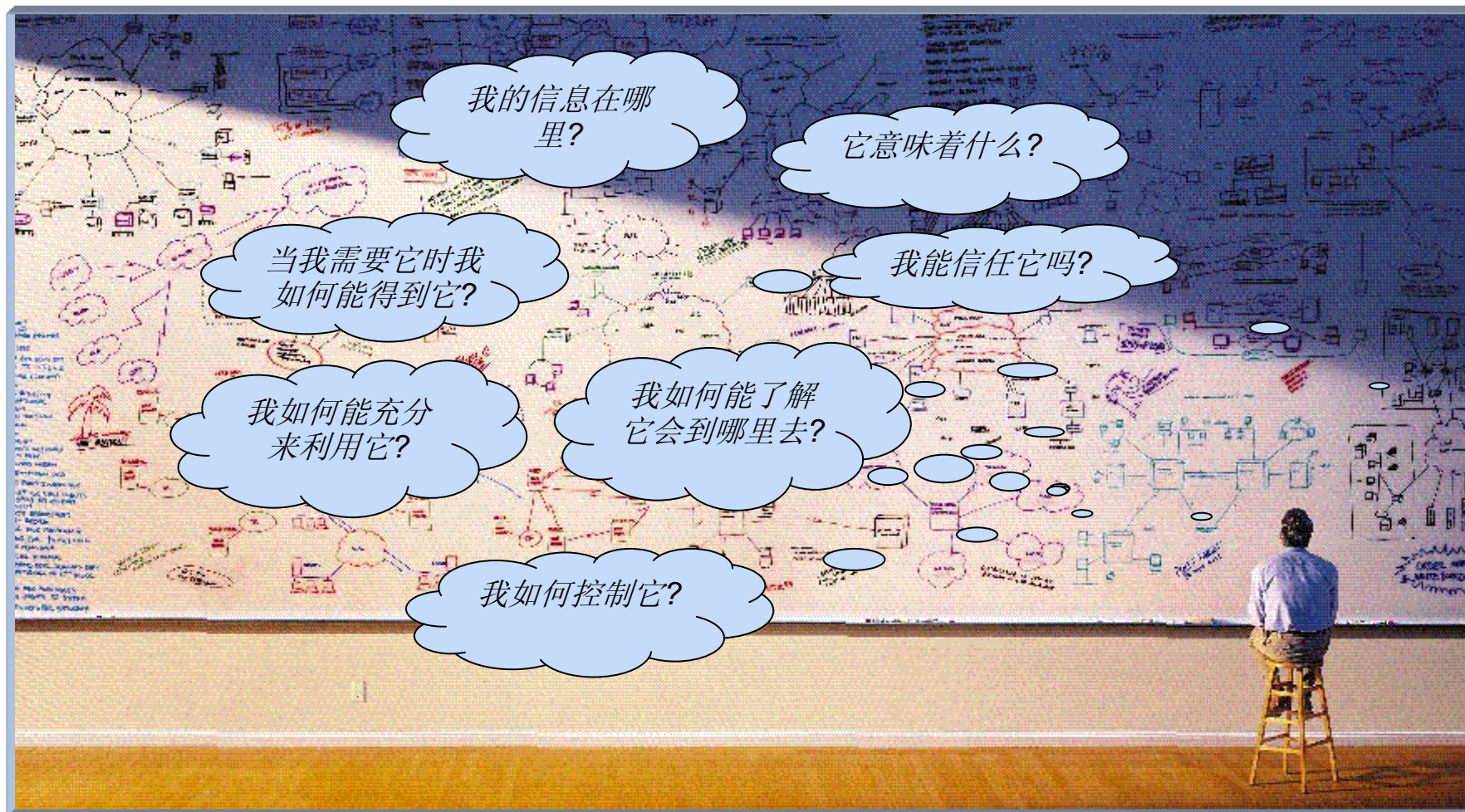


业务元数据的重要性





业务元数据帮助回答关键的问题



信息管理的挑战

业务用户

- ◆ 信息的真实性
- ◆ 信息孤岛
- ◆ 跨不同组织架构的信息无法共享
- ◆ 提供合适的信息给合适的人
- ◆ 信息的可用性
- ◆ 使用信息作为一种竞争的优势

IT用户

- ◆ 不同的平台,格式, 应用和数据库技术
- ◆ 数据和定义的不一致性
- ◆ 数据和系统的管理
- ◆ 对数据分析和信息的不断增加的需求
- ◆ 控制问题 **vs** 业务需求





IBM 2010 新锐洞察高峰论坛



我如何能知道我有一个精确的视图?



Business Users



Subject Matter Experts

我要想使用工具来帮助我的工作..



Developers

我如何能积极的与开发者协同工作?

简化管理, 部署和维护

如何实现数据治理, 安全和可扩展性?



Architects



DBAs

为什么我的工具不能更集成化?



Data Analysts



主要的开发主题



Business Users



Subject Matter Experts



Architects



Data Analysts



Developers



DBAs

- 简化系统开发的复杂性
- 简化系统集成的复杂性
- 实现业务和**IT**的有效的协同工作
- 使工具能更好的支撑我的工作
- 充分利用我已有的**IT**投资
- 在项目和企业的范围内提升重用性



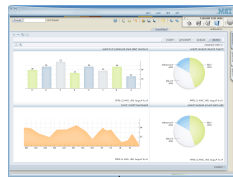


增强的协同性

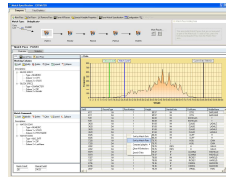
Role-based Tools with Integrated Metadata



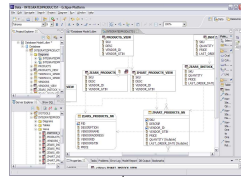
Business Users



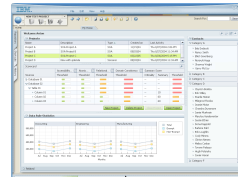
Subject Matter Experts



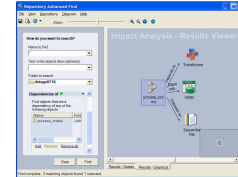
Architects



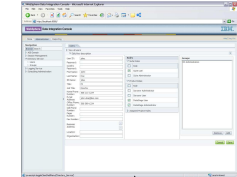
Data Analysts



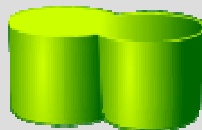
Developers



DBAs



Unified Metadata Management



Design Operational

- Simplify Integration
- Increase trust and confidence in information
- Facilitate change management & reuse
- Increase compliance to standards



他们用元数据做什么？

我们用它来发现并证实发生了什么？

我们用它来管理改变发生后的冲突分析



Business Users



Subject Matter Experts



Architects



Data Analysts



Developers



DBAs

我们用它来记录并连接业务概念

我们用它来理解并记录信息源头

我们用它来加速信息集成的设计

我们用它来加速信息集成的开发



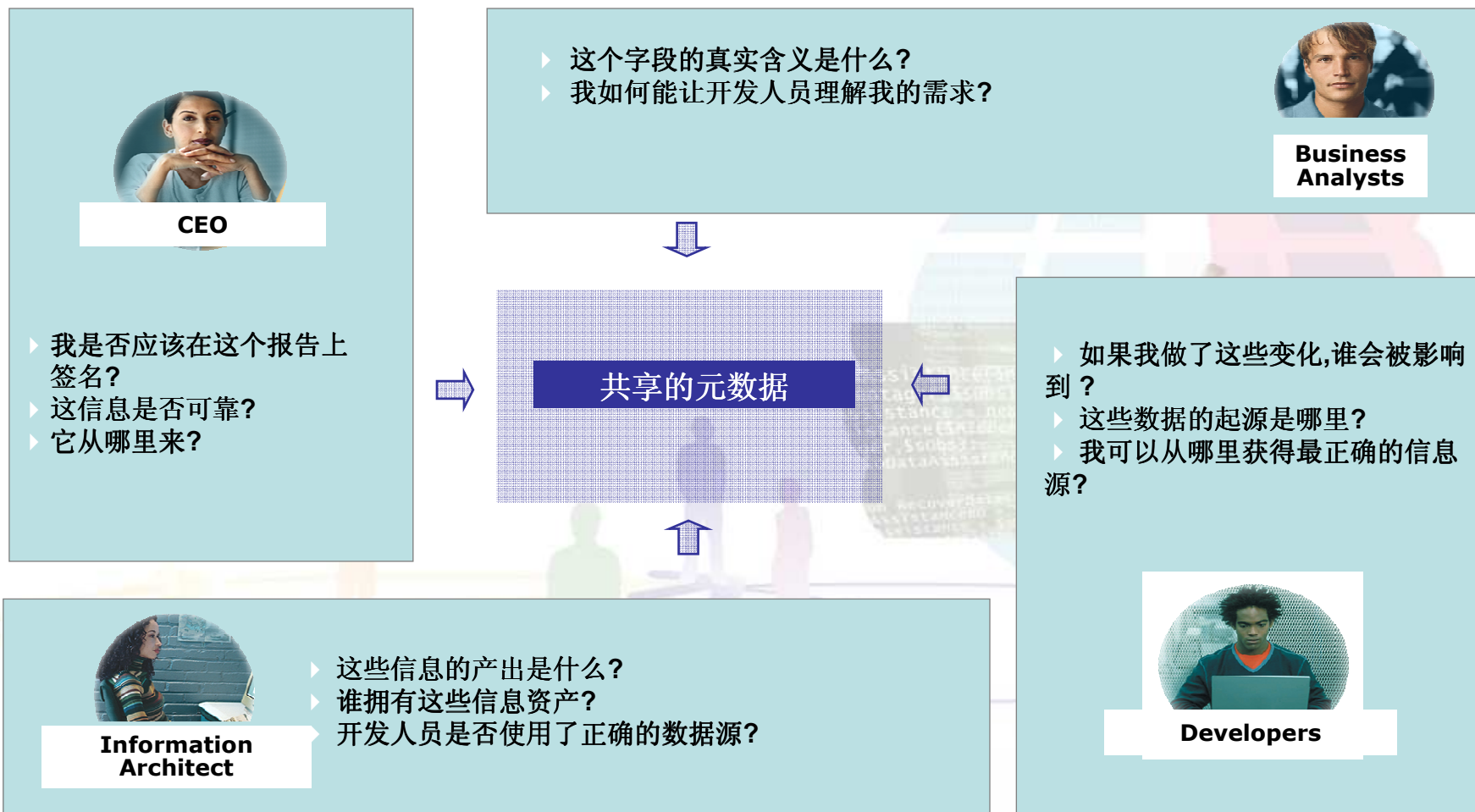


好的元数据带来的好处

- 元数据是一个手段,并不是终结
- 有效的使用元数据可提升用户的体验
 - “正确”的元数据
 - 给“正确”人
 - 在“正确”的地方
 - 在“正确”时间
- 为缺乏经验和非技术用户
 - 当信息集成是有业务用户来驱动时,它变得更为关键
- 最好通过SOA来实施
 - 技术逻辑和拥护界面的分离 **Separation of technical logic and UI**
 - 复杂逻辑的实现,如.:冲突分析
 - 使客户端较轻量

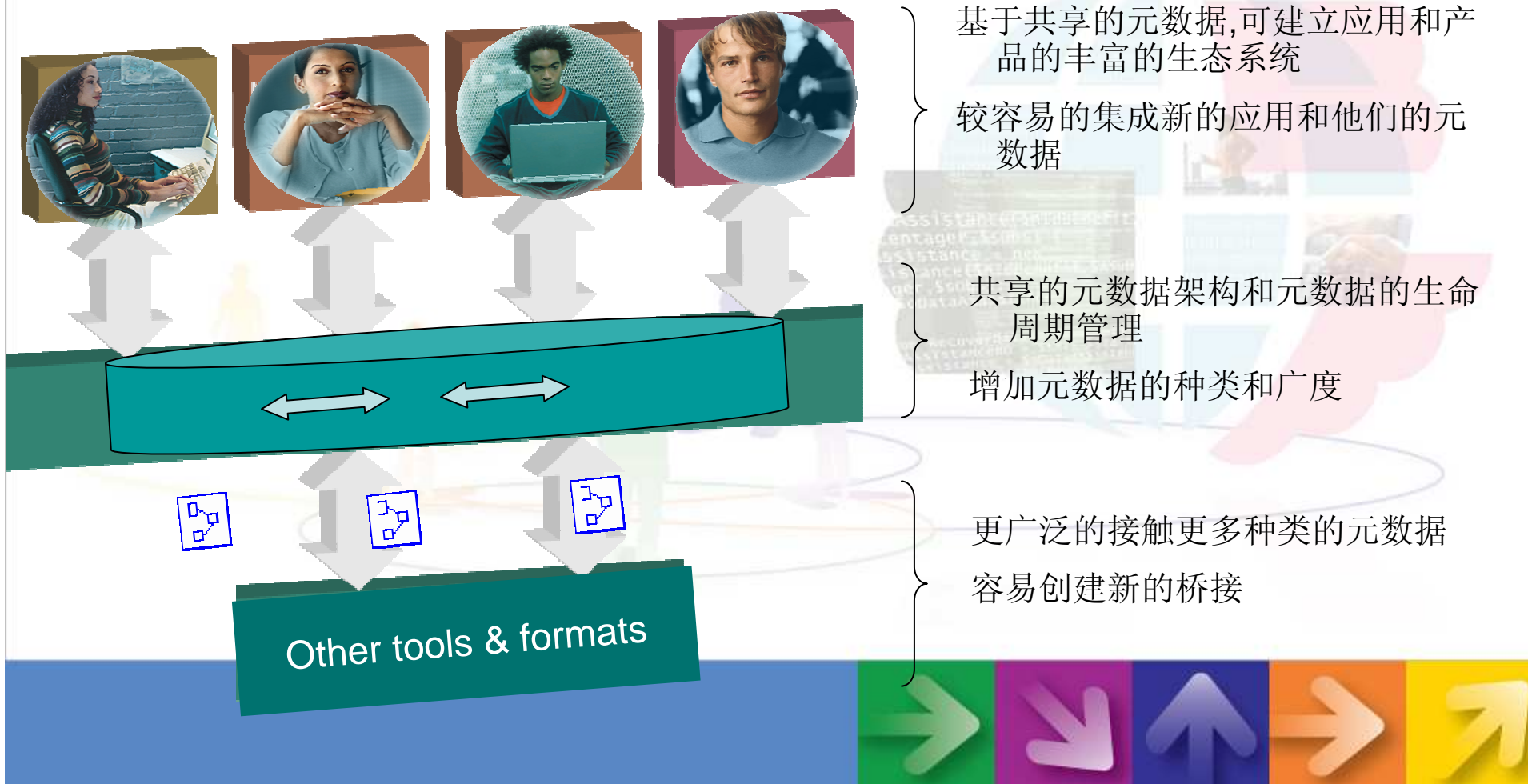


谁可以从好的元数据获益？



带来的价值

- 跨工具的集成的信息流 - 消除孤岛
- 方便改变,管理和重用
- 使标准增加一致性
- 增加信息的信任和自信





IBM 业务元数据管理的解决方案





业务用户



主题专家



架构师



数据分析师



开发人员



DBAs

IBM Information Server

Information Services Director

为整合信息和访问发布SOA服务

Business Glossary

归档业务术语 & 连接到数据源

QualityStage

标准化、合并和纠正信息

DataStage

组合和重构信息以用于新的用途

Federation Server

异构信息的虚拟化访问

Information Analyzer

分析 & 理解 源数据

Change Data Capture

数据增量获取

Metadata Server / Metadata Workbench

跨信息整合生命周期的统一的元数据管理

并行处理

丰富的应用、数据和内容连接的支持



业务元数据

用Business Glossary来管理和展示企业的业务元数据

- 基于Web的业务元数据管理工具，提供业务元数据的编辑，共享和管理功能。
- 通过提供元数据的业务含义，可以将技术元数据和业务元数据有效结合起来，从而提供完整的元数据视图
- 可以提供多种业务元数据管理，例如元数据所有者，术语和分类的管理
- 业务元数据可以通过XML批量导入或者通过Business Glossary前端界面插入到Repository中

- 指标表
- 维度表
- 指标维度对应表



Business

- 普通业务人员
- 领导
- 应用系统开发人员
- ETL开发人员
- 业务元数据维护人员



Subject Matter Experts



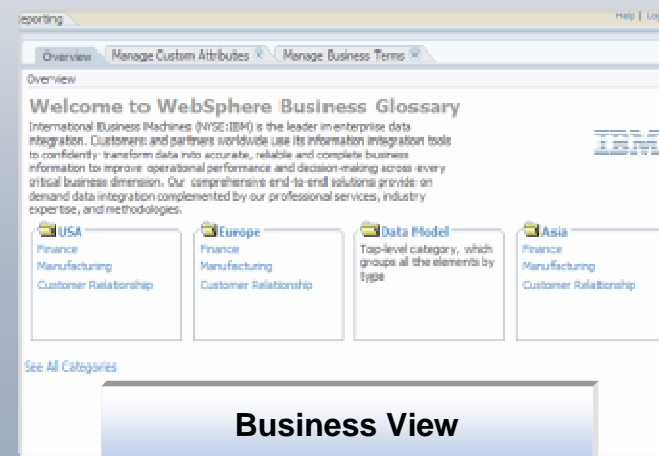
Business Users

Understand



WebSphere Business Glossary

Create and manage business vocabulary and relationships, while linking to physical sources





业务术语

- 编辑和管理标准的业务术语，包括例子，缩写，所有者，同义词等等。
- 可以将这些业务术语同在**metadata repository**中的技术元数据结合起来从而提供完整的元数据视图
- 在用户的自然语言和在**metadata repository**高度结构化的信息之间，提供了一层语义解释层。
- 可以从**XML**输入业务术语

例如：

Customer Number – 十位数字的客户编码，有时又称为Customer ID，样值的形式为：1023479658.

所有者: 销售部, 李军

同义词: Customer ID

参见: Customer Type



业务分类

- 建立和管理业务层次
- 用以对业务术语归类
- 定义了 **Business Glossary** 中的浏览结构，同时构建了查询结果
- 可以从 **XML** 输入

业务考核

- 指标

- 业务部门

- 子部门

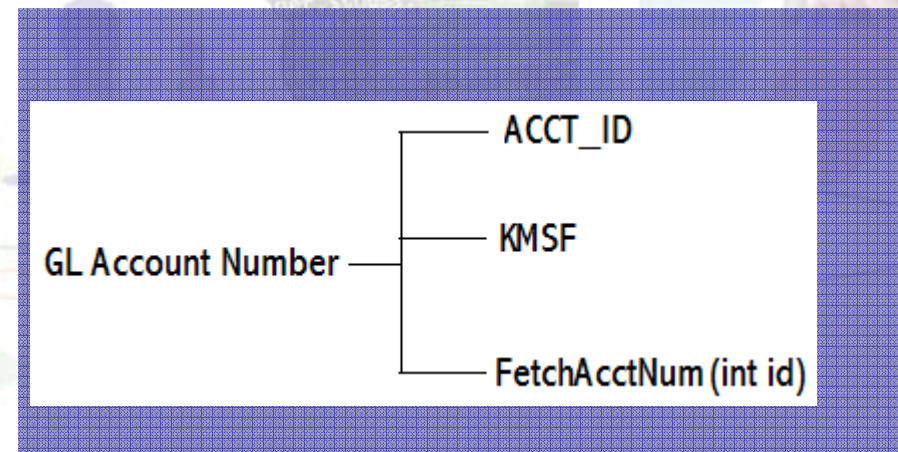
业务元数据的协同管理

- 可以对**glossary**中的元数据对象增加注释和描述
- 可以使业务用户丰富元数据的业务含义
- 对业务元数据的注释一经修改，可以立刻共享给其他用户
- 对业务元数据的注释和描述可以由管理员监控，同时可以反馈给系统管理员

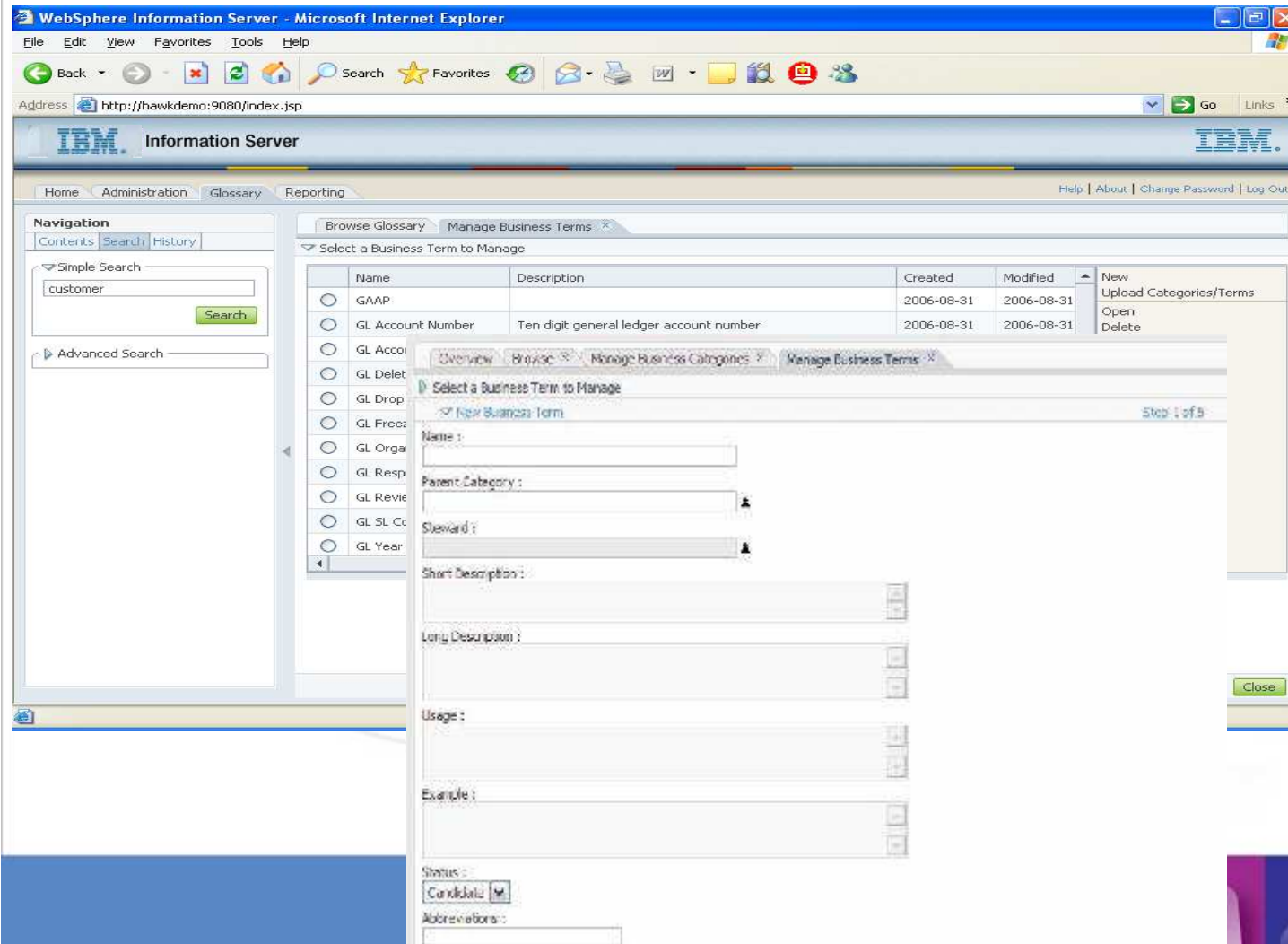
我发现总帐机构的最后两位代表的是子部门，但通常是空的。

联系并下钻到技术元数据

- 业务术语可以用以描述其他应用的元数据
- 可以下钻到技术元数据
 - Data Models
 - ETL Flows
 - Business Intelligence
- 可以浏览这些技术元数据
 - View all attributes
 - View all relationships
 - Set or view stewardship
 - Run analysis
 - View or add Notes

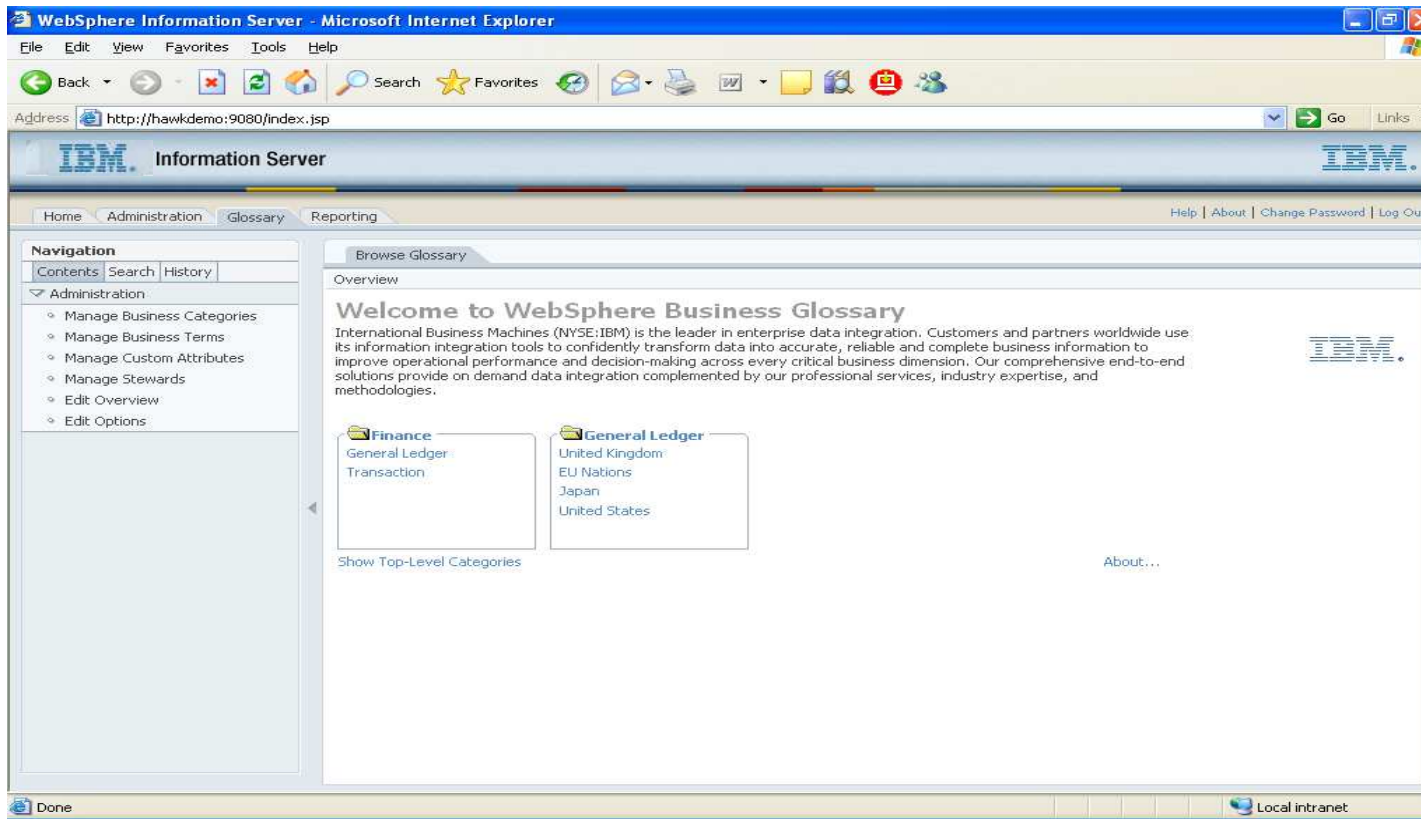


业务元数据的管理工具



- 基于**Web**的存取和协同管理业务元数据
- 可以使所有业务用户协同处理，例如业务元数据的定义，同义词，样例和分类
- 提供丰富的元数据管理，包括报表，查询，影响分析和注释等

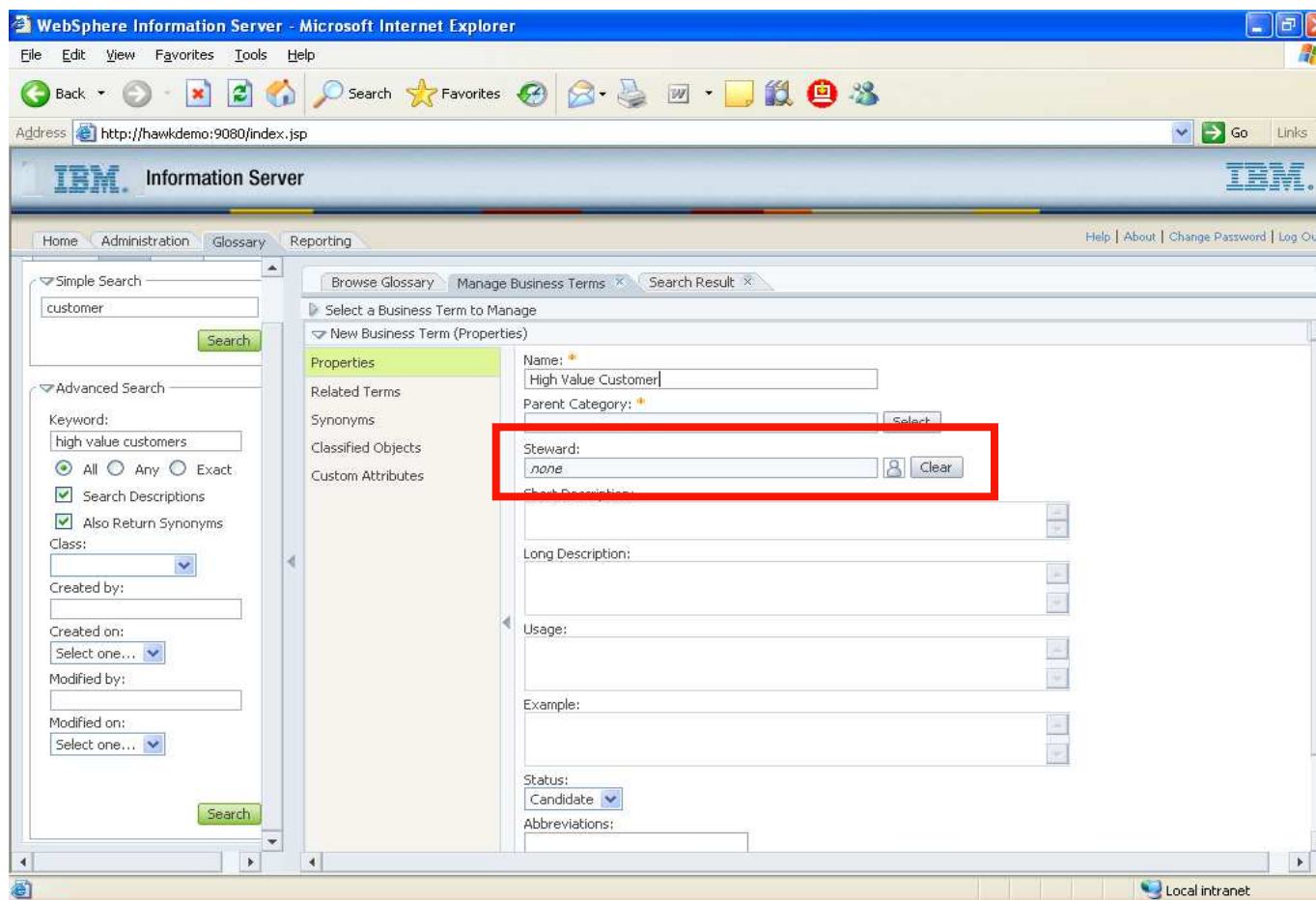
业务和IT的协同



- 业务术语可以归类和管理不同的机构
- 在共享的 **Repository** 的业务元数据可以进行主动的管理
- 可以对业务术语和分类增加注释
- 可以将业务术语联系到表和字段



管理元数据的所有者



- 允许指派业务术语的所有者
- 允许业务术语的所有者协同管理注释和描述
- 支持候选的业务术语 和提升
- 所有者管理标准的业务术语定义





IBM WebSphere Business Glossary

价值

- ▶ 提供了业务元数据的管理
- ▶ 可以捕捉业务视图
- ▶ 紧密协同业务和IT
- ▶ 提供了业务框架，有助于业务决策



好处

- ▶ 缩短了分析时间
- ▶ 帮助IT部门了解业务目标
- ▶ 提供了不断发展的业务元数据视图
- ▶ 加快了业务规则的开发



业务元数据在桌面的展示

用 **Business Glossary Anywhere** 在任何桌面应用上展示业务元数据

- 支持Popup 桌面的搜索和查询
- 可以应用到多数桌面应用
- 增强了Business Glossary的查询
- 支持tighter collaboration



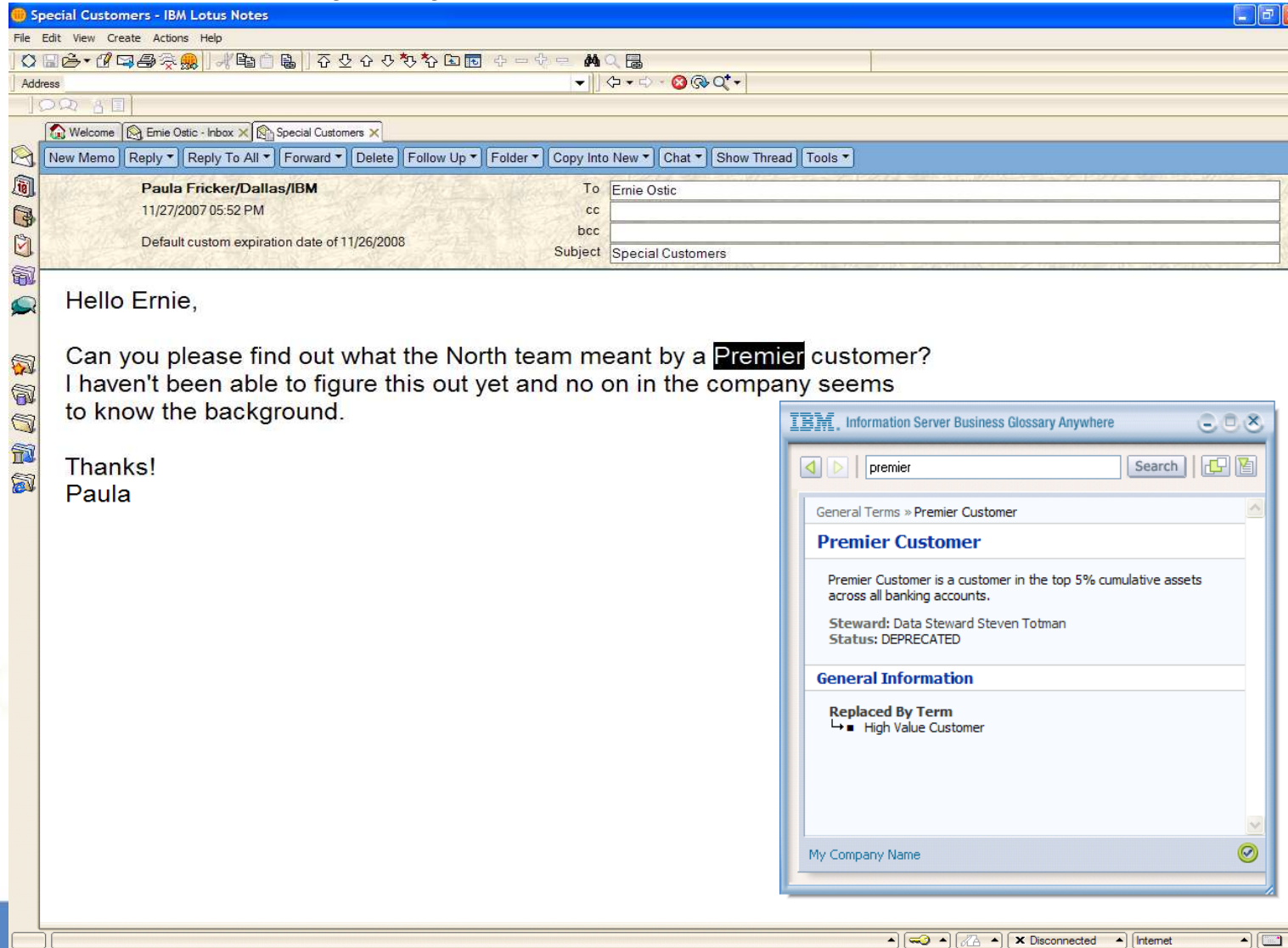
跳出窗口,展示业务元数据

点击深色部分

Subcategory	Revenue	Forecast	Units Sold
Electronics	\$ 13,822,698		31,4
Miscellaneous	\$ 9,745,547		27,4
Audio Equipment	\$ 9,011,792		31,2
TV's	\$ 8,293,859		26,8
Computers	\$ 3,510,964		27,2
Average	\$ 9,231,355		28,4
Total	\$ 55,388,128		170,8



Business Glossary Anywhere with Lotus Notes



The screenshot shows a Lotus Notes application window titled "Special Customers - IBM Lotus Notes". The interface includes a menu bar (File, Edit, View, Create, Actions, Help), a toolbar, and a navigation pane on the left. The main content area displays an email from Paula Fricker/Dallas/IBM to Ernie Ostic, dated 11/27/2007 05:52 PM. The email body contains the following text:

Hello Ernie,

Can you please find out what the North team meant by a **Premier** customer? I haven't been able to figure this out yet and no one in the company seems to know the background.

Thanks!
Paula

An IBM Information Server Business Glossary Anywhere window is overlaid on the email content. It shows a search for "premier" with the following results:

- General Terms » Premier Customer
- Premier Customer**
- Premier Customer is a customer in the top 5% cumulative assets across all banking accounts.
- Steward: Data Steward Steven Totman
- Status: DEPRECATED
- General Information**
- Replaced By Term
 - High Value Customer

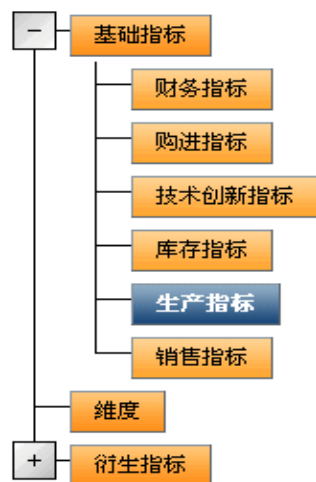
The window also includes a search bar, navigation arrows, and a "My Company Name" field at the bottom.



客户案例：上海烟草

上海烟草（集团）公司数据中心元数据管理平台

• Search • Category Tree • About • Help



Expand All

CATEGORY DETAILS:

生产指标

TERMS:

- 卷烟产量





使用**InfoSphere Business Glossary** 来管理你的业务元数据的好处





好处:在业务人员和技术人员之间创建共同的字典

Database = DB2

Schema = NAACCT

Table = DLYTRANS

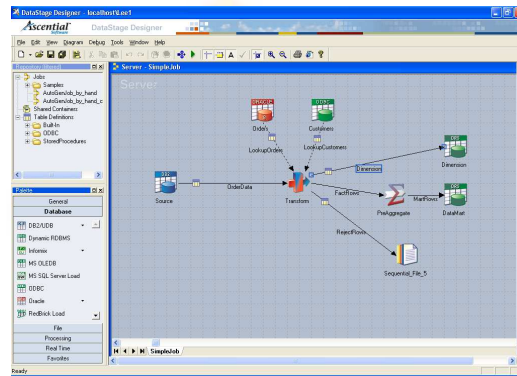
Column = ACCT_NO

data type = char(11)

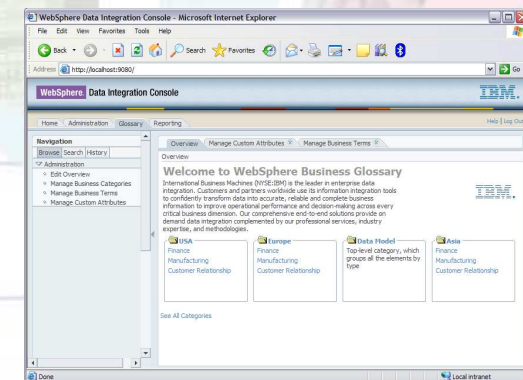


GL Account Number

The ten digit account number. Sometimes referred to as the account ID. This value is of the form L-FIIIIVVVV.



InfoSphere DataStage



InfoSphere Business Glossary



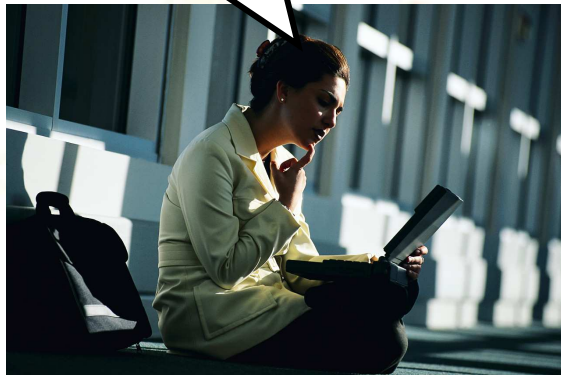
好处: 协同工作, 共享经验、想法和意见

I've noticed that the last two digits of the GL Organizational Unit, which indicate the sub-department, are often blank.



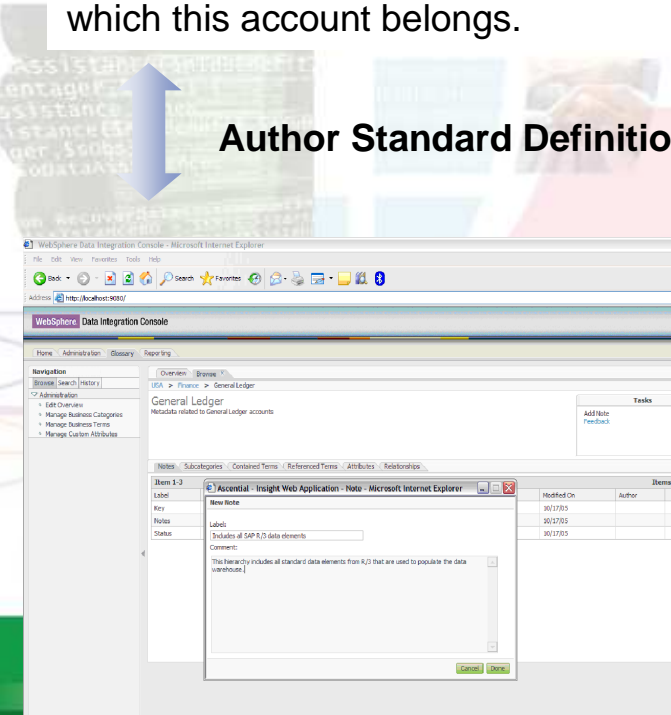
GL Organizational Unit

STEWARD: Controllers Office
FORMAT: X(7)
DEFINITION: A seven digit number designating the organizational unit to which this account belongs.



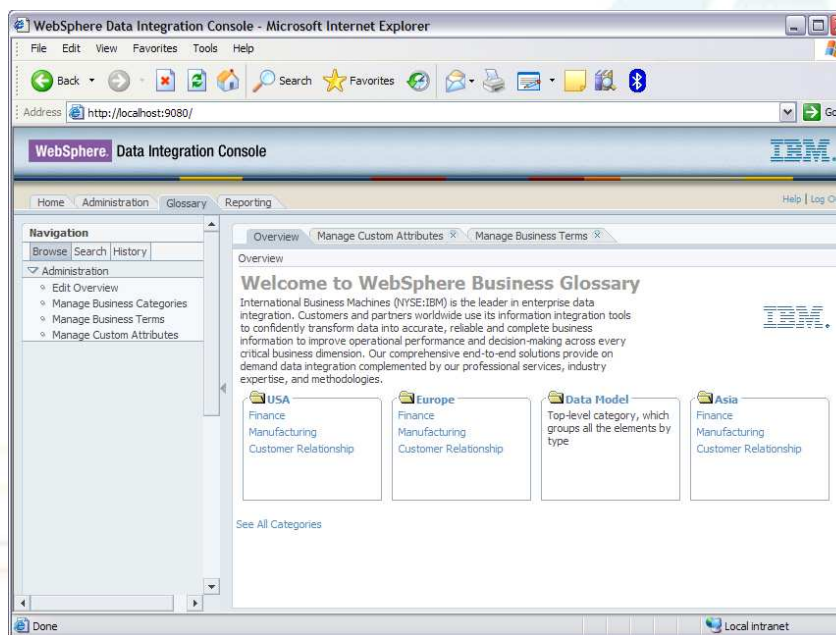
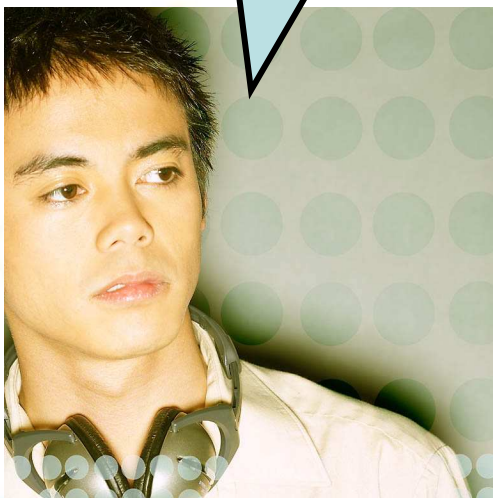
Annotate Any Object

Author Standard Definitions



好处: 能快速的访问关键的业务元数据

我不需要特殊的培训就能使用 business glossary. 他的功能正如我希望的那样.

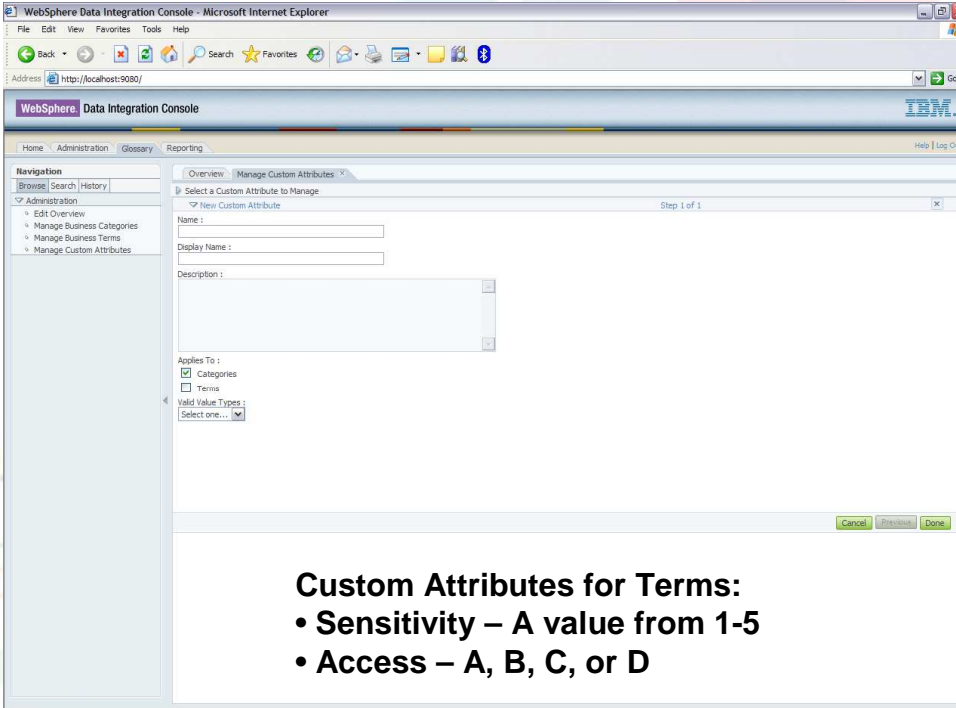


我可以很快把它建好, 使我的人们能立即看到这些常用的业务元数据和它的层次化



好处: 客户化应用来满足你用户的特殊需要

可创建每一个业务元数据
共有客户化的字段属性



WebSphere Data Integration Console - Microsoft Internet Explorer

Address: http://localhost:9080/

WebSphere Data Integration Console

Home Administration Glossary Reporting

Navigation: Administration > Manage Custom Attributes

Select a Custom Attribute to Manage

Name:

Display Name:

Description:

Applies To: Categories Terms

Valid Value Types:

Cancel Previous Done

Custom Attributes for Terms:

- Sensitivity – A value from 1-5
- Access – A, B, C, or D

Done





Thank You

