



Complete your Data Lineage...

with IBM and Compact Solutions



Siddhartha Banerjee PMP

Sales Leader - Americas & Sr Principal

Sid.Banerjee@compactbi.com (m) +1.862.579.7089 www.compactbi.com

Innovative software and services for ETL Migrations, Metadata Management,
Automated Batch QA, and NoSQL data analytics for Big Data environments

Agenda

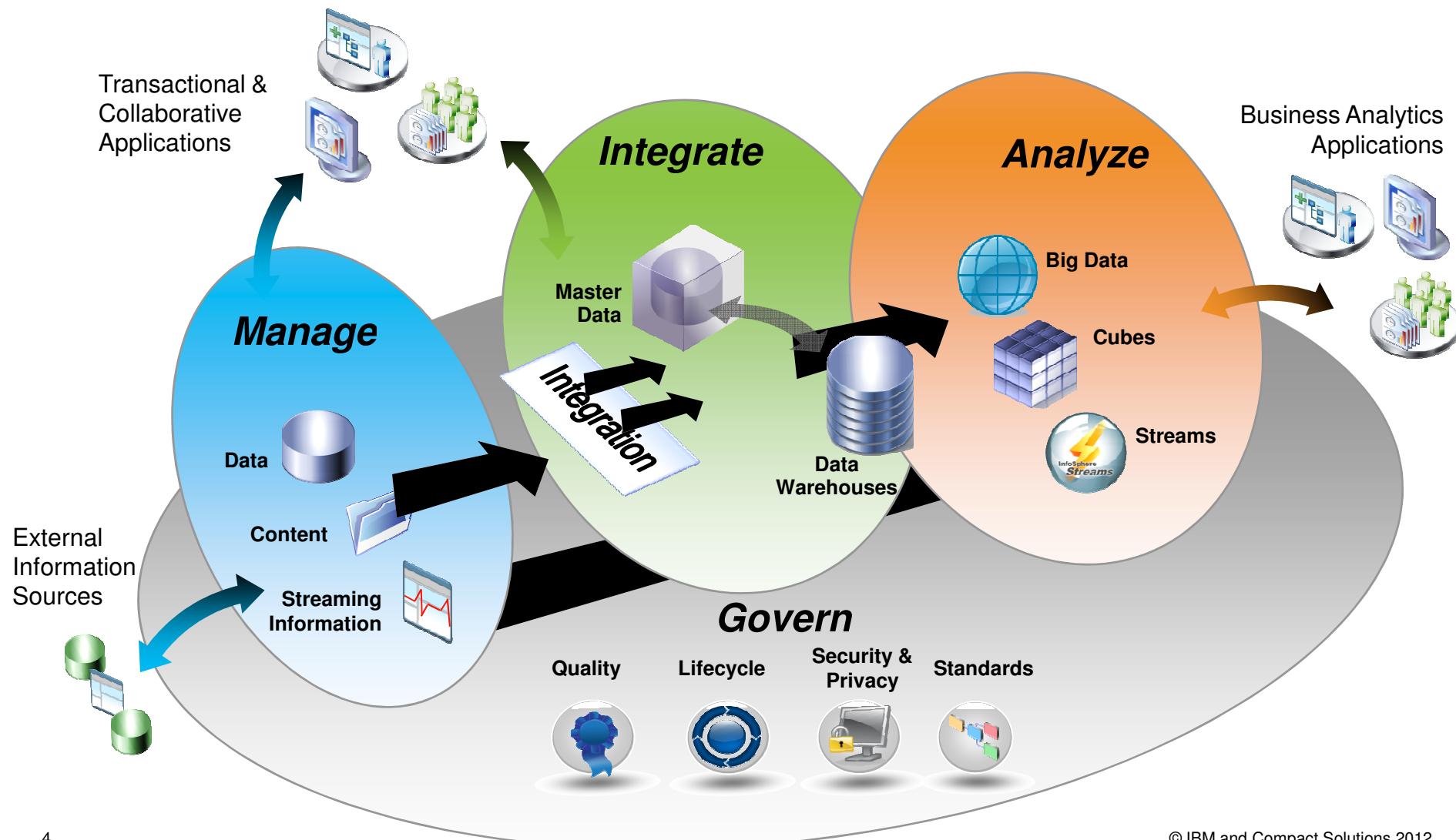
- Need for integrated metadata management
- Recap IBM InfoSphere metadata solution
 - Business Glossary
 - Metadata Workbench
- Compact MetaDex
- Demo
- Case Study
- Question & Answer

Compact Solutions Overview

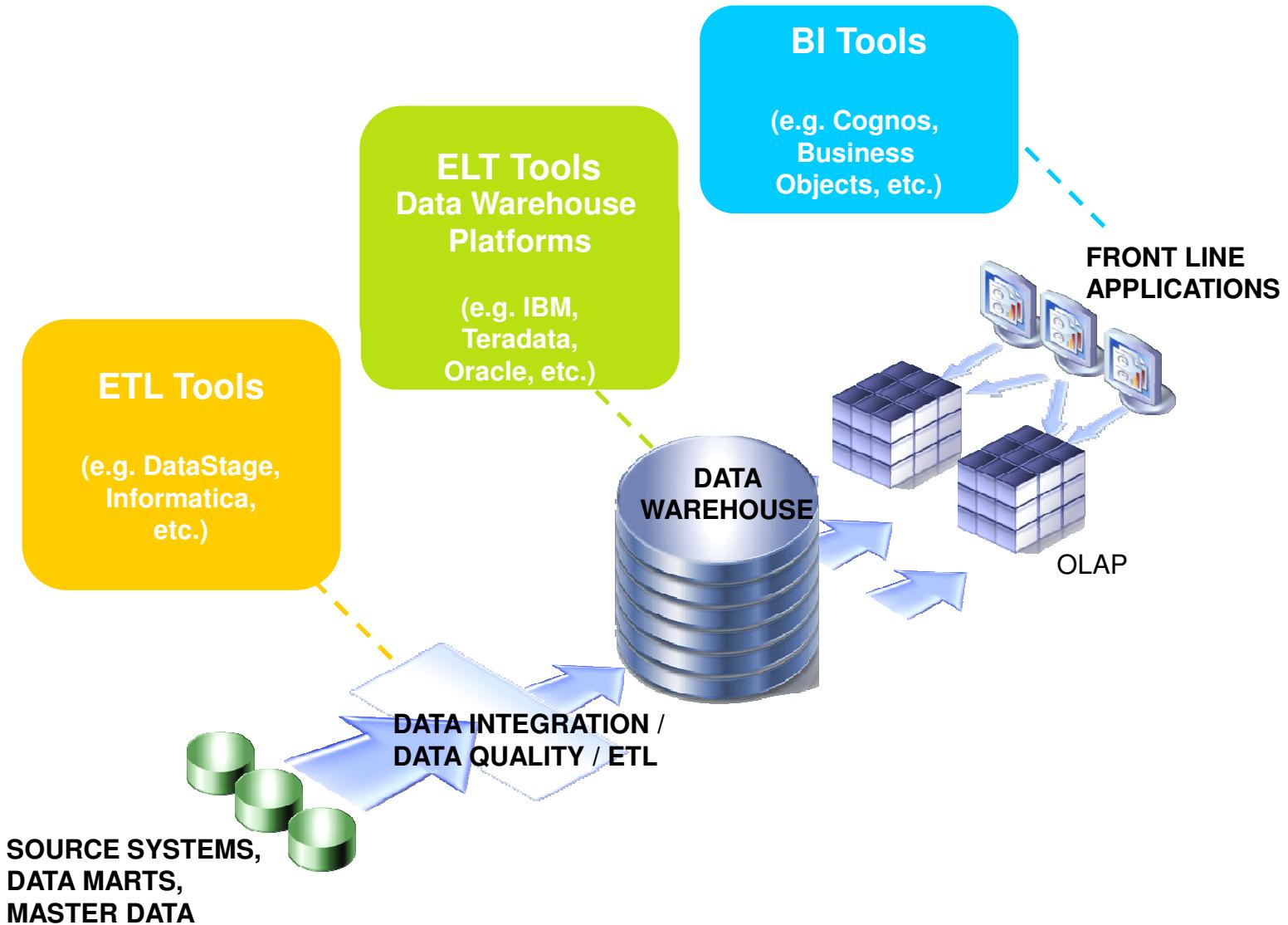
- **High-performance, enterprise-class data management solutions**
- **We provide**
 - Solutions (Metadata Integration, ETL migrations, Batch QA Automation, Big data analytics,...)
 - Technology expertise (IBM InfoSphere, Informatica, Ab Initio, Teradata, etc)
 - Best practices for full life cycle of data integration/management
- **Presence in**
 - Chicago (US) – Worldwide HQ
 - Krakow (Poland) – Innovation labs
 - Ahmedabad (India) - Dual shore services
- **Industry recognition by Forrester as tech solution provider**
- **Marquee customers in US and Europe**
 - Electronic Arts, Sears, United Airlines, Shutterfly, Sony Pictures, Mobistar, BNP Paribas, Honda, Orange Telecom
- **IBM InfoSphere Portfolio Premier Partner**



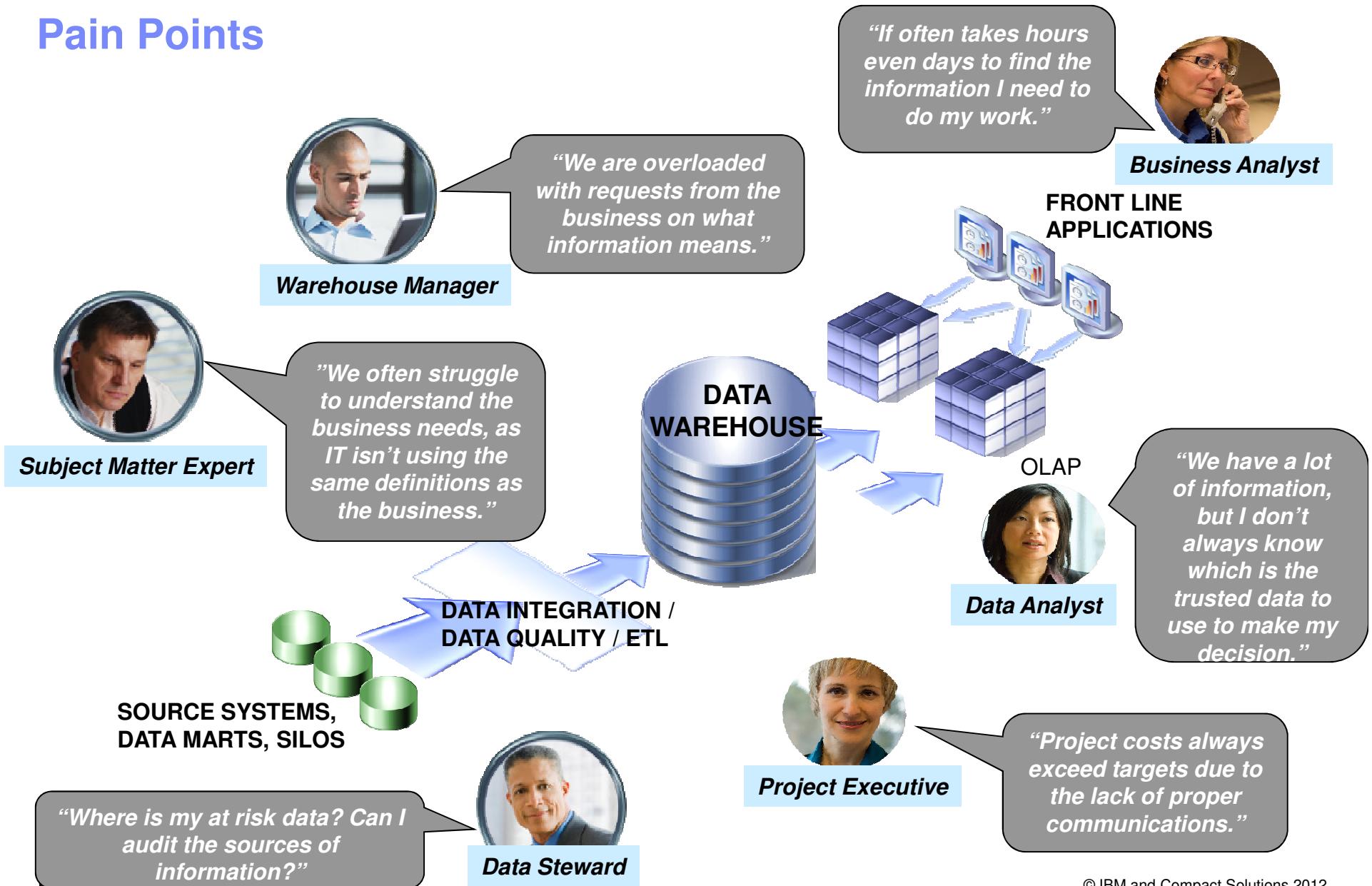
Information Management Supply Chain



Information Management Logical Architecture



Pain Points



Challenges of Every Data Integration Project

- Lack of communication and collaboration across teams
- Inability to find information quickly
- Information ambiguity and lack of full understanding
- Lack of confidence and trust in information
- Poor data governance



Why Metadata is Important for Business Managers?

"What are the business drivers for metadata management?"



Base: 56 IT professionals with a metadata strategy (multiple responses accepted)

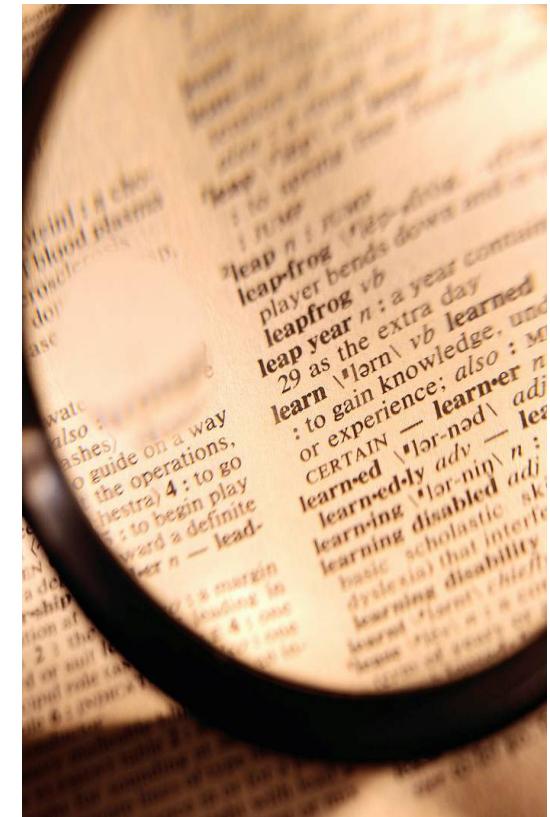
Source: November 2009 Global Information Architecture Online Survey

["Enterprise Architecture Must Lead Enterprise Metadata Management Strategies," Forrester Research, Inc., January, 2011](#)

Gartner's Recommended Best Practices

To avoid terminological debates, confusion, information inconsistency, **Gartner recommends:**

- “Establish a common vocabulary for discussing information consistency issues
- Use concrete examples to gain agreement when in doubt about terminology
- Drive practitioners of the various information management disciplines to resolve, in detail, how they will work together on implementation activities”



Gartner 2 July 2009, A Master Vocabulary for Ensuring Consistency Across Information Management Disciplines
ID Number: G00165576

What is Metadata?

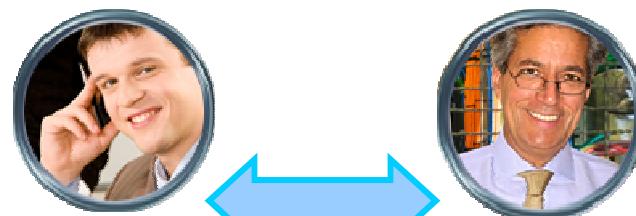
Business Metadata

- Business terminology
- Business rules definitions
- KPI and metric definitions
- Link to supporting technical assets
- Business lineage

Technical Metadata

- Source and target data definitions
- BI model and report definitions
- ETL jobs, stored procedures
- Related terms
- Data lineage

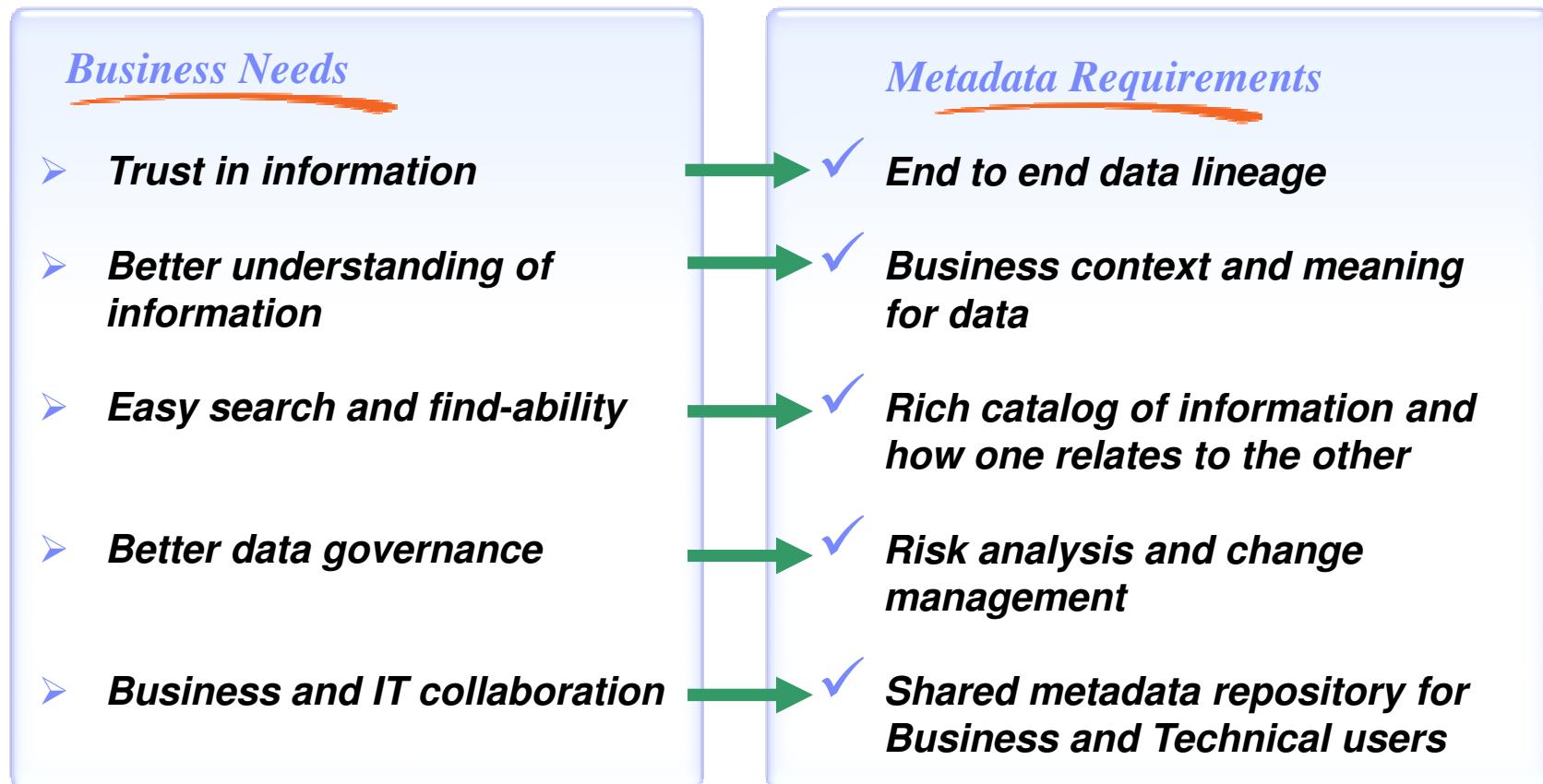
Category: Costs
Term: Tax Expense
Description: Tax to be paid on Gross Income
Status: Current
Data Steward: Roger West
Related Term: Tax Liability
Related Asset: Income_tbl



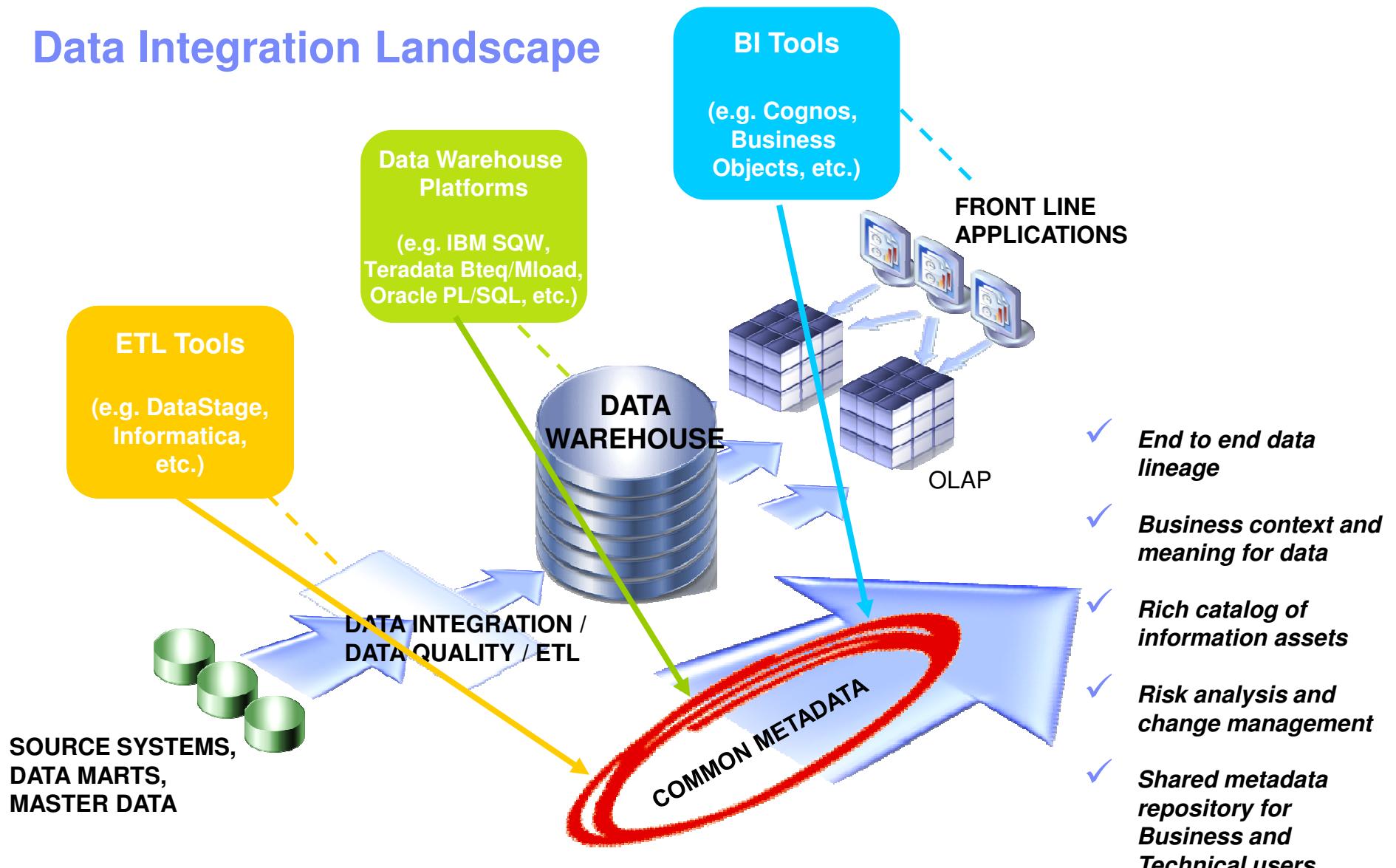
Business & Technical teams need to collaborate

Source: Income_tbl (Oracle Table)
Derivation: SUM(TAXAMT)
Data Rule: TAXEXP > \$1
ETL Job: TXExp
Business Term: Tax Expense

Metadata Requirements in Data Integration Projects



Data Integration Landscape



IBM Solutions for Metadata

Category: Costs

Term: Tax Expense

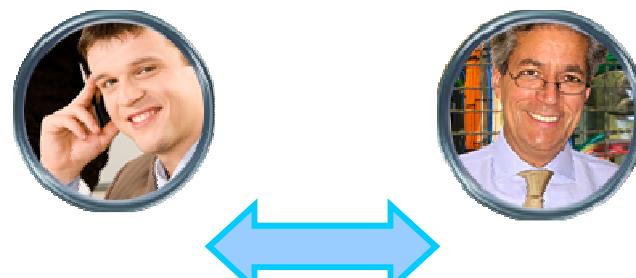
Description: Tax to be paid on Gross Income

Status: Current

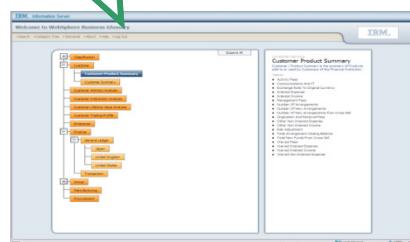
Data Steward: Roger West

Related Term: Tax Liability

Related Asset: Income_tbl

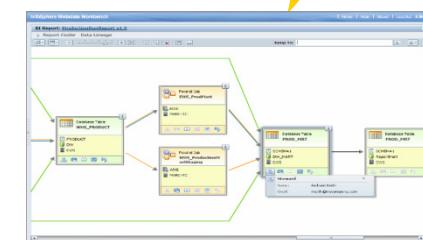


*Collaboration
between Business
& Technical Teams*



Business Glossary

Common
Metadata
Repository



Metadata Workbench

Source: Income_tbl (Oracle Table)

Derivation: SUM(TAXAMT)

Data Rule: TAXEXP > \$1

ETL Job: TXExp

Business Term: Tax Expense

InfoSphere Business Glossary



Business Glossary

Create and manage business vocabulary and relationships and related to physical sources

WebSphere Business Glossary

Welcome to WebSphere Business Glossary

Category Details: Aggregate Measures

TERMS:

- Accumulated Other Comprehensive Income
- Activity Fee Income
- Actual Loss Amount
- Annualized Average Risk Weighted Assets Half
- Annualized Average Risk Weighted Assets Quarter
- Annualized Average Risk Weighted Assets Year
- Annualized Average Unweighted Exposure After Credit Risk Mitigation Technique H
- Annualized Average Unweighted Exposure After Credit Risk Mitigation Technique M
- Annualized Average Unweighted Exposure After Credit Risk Mitigation Technique Q
- Annualized Average Unweighted Exposure After Credit Risk Mitigation Technique Y
- Annualized Average Unweighted Exposure Before Credit Risk Mitigation Technique H
- Annualized Average Unweighted Exposure Before Credit Risk Mitigation Technique M
- Annualized Average Unweighted Exposure Before Credit Risk Mitigation Technique Q
- Annualized Average Unweighted Exposure Before Credit Risk Mitigation Technique Y

Requirements

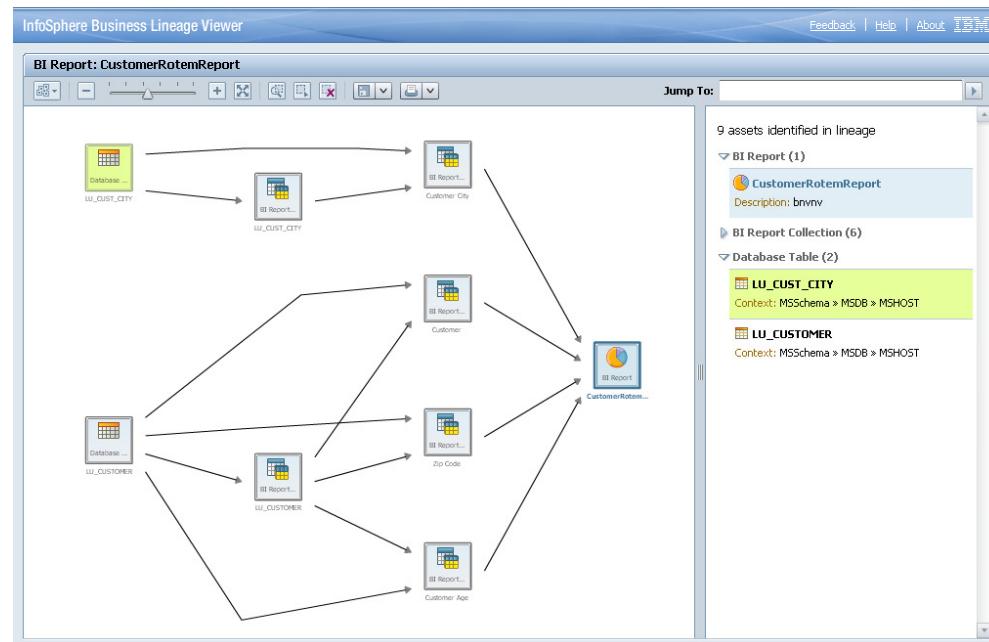
- Capture business terms and classifications
- Link business terms and classifications to IT assets
- Identify data stewards and make glossary accessible

Benefits

- Context for information is available to everyone, immediately
- IT projects are aligned with data governance
- Collaboration increases across business and IT

Business Governance – Trusted Sources of Information

- Empower the “business user” to make critical decisions with trust and confidence in information
- Understand quickly where information originates
- Reduce cycles spent across teams validating data accuracy



Simple to understand and use with no training required!

*Includes only the **critical path of information** which the business user cares about.*

Business Context for Every Heterogeneous IT Asset

- Starting point for designing information governance projects
- A “Business Users” gateway into Heterogeneous IT environment
 - ✓ *Data Sources*
 - ✓ *ETL Jobs*
 - ✓ *Stored Procedures*
 - ✓ *Applications*
 - ✓ *BI Reports*
 - ✓ *Business Processes*
 - ✓ *Data Models*
 - ✓ *Web Services*
 - ✓ *More...*

InfoSphere Business Glossary Browser : Welcome to InfoSphere Business Glossary

Help | About | Log Out IBM.

Search | Category Tree | Browse ▾ | Viewing Permissions

Term Details [Print](#) [Feedback](#)

Finance > General Ledger > United States >

GL Account Number

Short Description: Ten digit general ledger account number

Long Description: The ten digit account number. Sometimes referred to as the account ID. This value is of the form L-FIIIIIVVV.

Status: Standard

General Information

Assigned Assets

- Customers_Table
- Marketing Campaign Email Message
- SimpleReport1

Notes

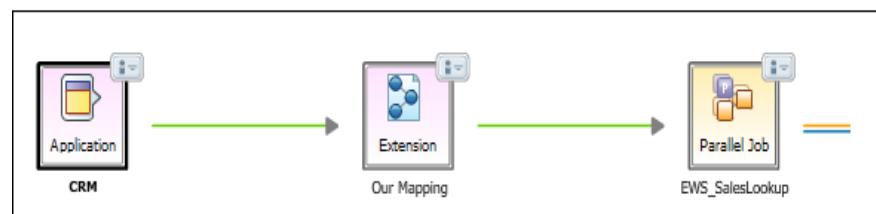
History

InfoSphere Metadata Workbench



Metadata Workbench

Support information governance with traceability on data movement, modeling & BI applications



Requirements

- Understand the impact of making changes to information environment
- Visualize and trace information flows across enterprise landscape
- Access and report on operational metadata

Benefits

- Avoid system disruptions
- Provide audit information for data governance
- Build trust with LOB users

Catalog and Report on All Metadata Relationships

Window into Your Data Integration World

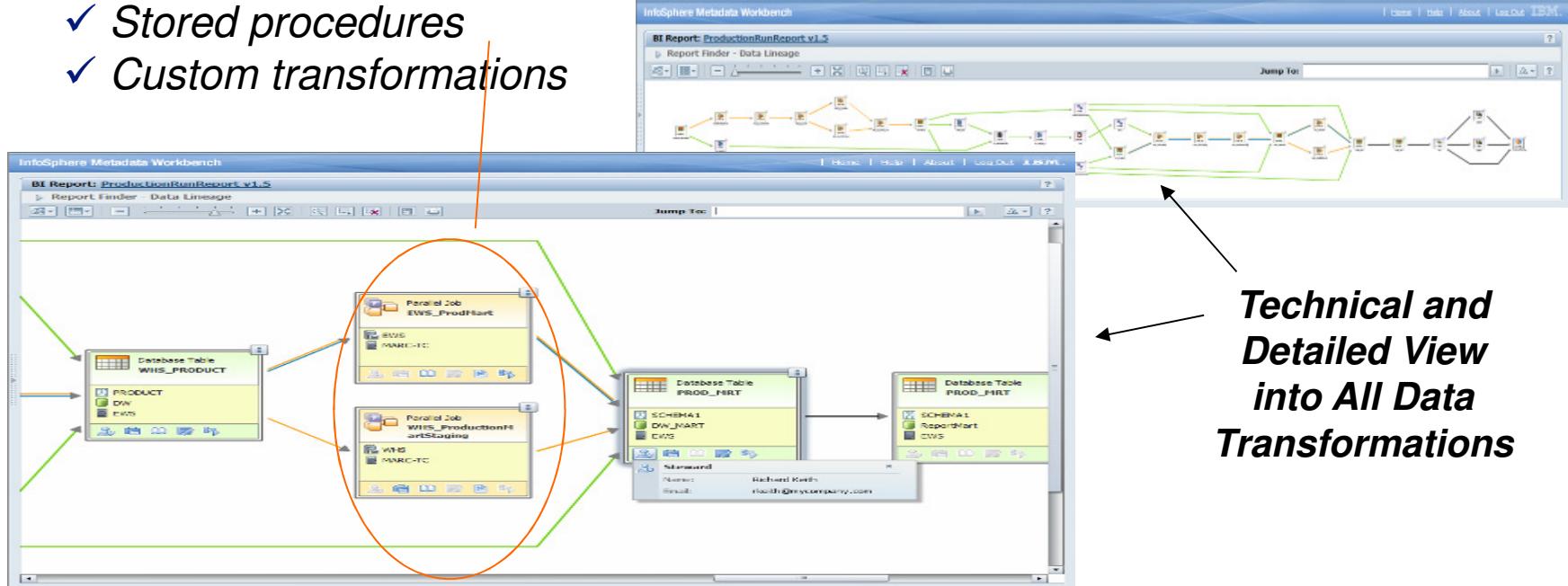
- ✓ Business Rules
- ✓ Quality Rules
- ✓ Business Terms and Definitions
- ✓ Data Stewards
- ✓ Runtime and Operational Statistics
- ✓ Source and Target Systems
- ✓ Table and Fields, Legacy Files
- ✓ BI Reports, Marts, & Warehouse Cubes
- ✓ Data Models, Logical Models
- ✓ ETL Jobs, FTP Processes, Scripts

| Database Design Usage Results (1-5 out of 5) | | | | | |
|--|--------------------------|---------------|------------------------|---|---|
| Tables | | Read by Job | | Written by Job | |
| Asset Name | Description | Steward | Schemas | Tables | Table Analysis Summary |
| BANKDATA | Bank Data Warehouse | Richard Keith | BANK BANK1 BANK2 | CUSTOMER BANKDEMOACCOUNTS BANKDEMOCHECKING BANKDEMOSAVINGS | CUSTOMER BANKDEMOACCOUNTS BANKDEMOCHECKING BANKDEMOSAVINGS |
| DW | Data Warehouse | Richard Keith | PRODUCT SALES | PLANT PRODUCTION WHS_PRODUCT WHS_SALES | WHS_PRODUCT WHS_SALES |
| DW_MART | Data Mart Reporting Data | Richard Keith | SCHEMA1 | PROD_MRT SALES_MRT | |
| Report Mart | | | SCHEMA1 | PROD_MRT SALES_MRT | |
| SALES | | | SCHEMA1 | SLS_LOOKUP | |

Information Governance

Visibility to Trusted Sources of Information

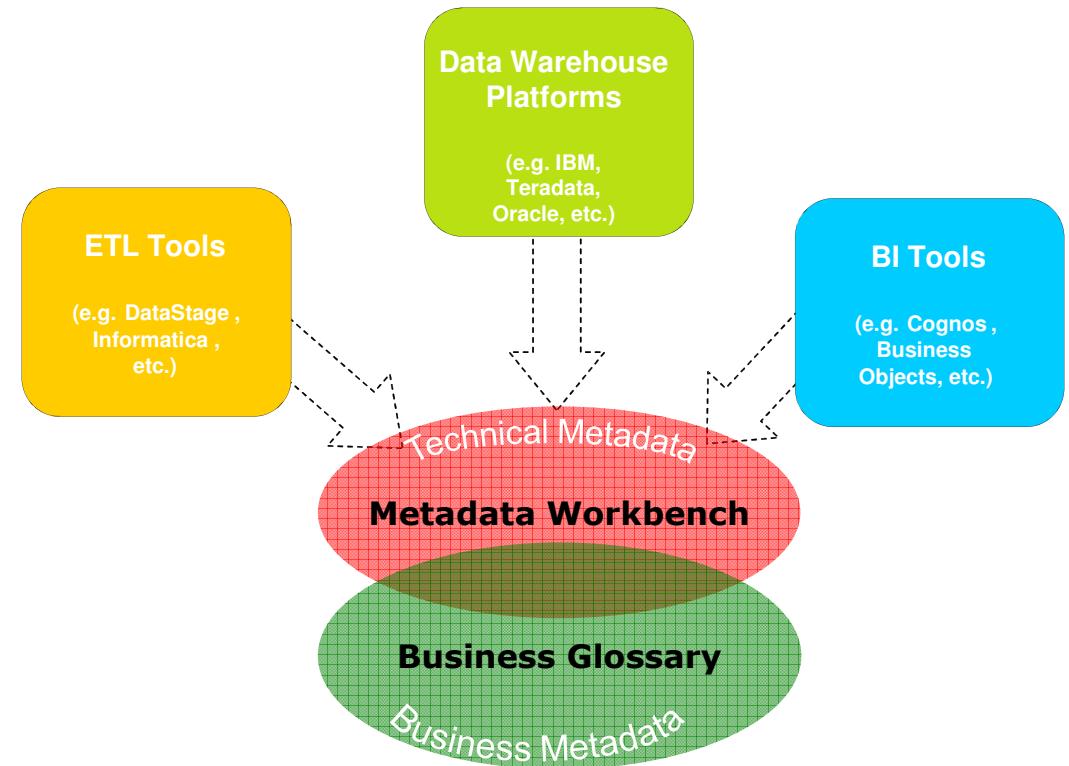
- **Extended visibility to Heterogeneous Data Integration Processes – including Information Server and 3rd Party**
- Popular business use cases
 - ✓ *3rd party ETL tools and applications*
 - ✓ *Mainframe COBOL programs*
 - ✓ *External scripts, Java programs, or web services*
 - ✓ *Stored procedures*
 - ✓ *Custom transformations*



Compact *MetaDex*

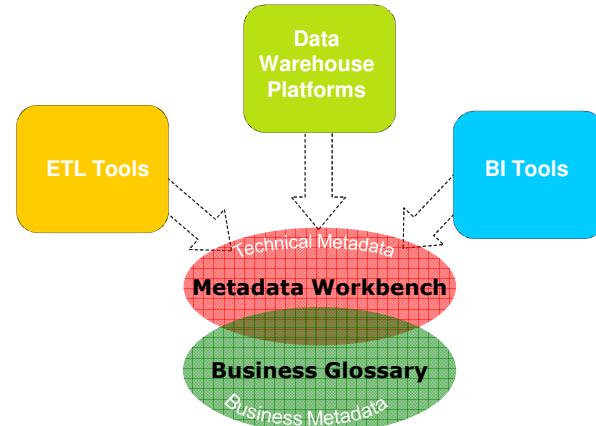
Embraces and Extends IBM InfoSphere Metadata Workbench

- A tool-accelerated services solution
- Integrates metadata from diverse data management processes
- Uses InfoSphere Metadata Workbench to consolidate heterogeneous metadata into IBM InfoSphere Platform



*The bridge to bring a multi-vendor data integration environment
into InfoSphere's single repository*

Compact *MetaDex* Solution Vision



Data Governance

- Data Lineage
- Data & Process Audits
- Business Rules



Process Improvement

- ETL overlaps
- Redundancies
- Gaps: Unused objects
- Technology standardization



Developer Productivity

- Process documentation
- Change Impact analysis
- Localize code changes
- Modular development



MetaDex DEMO

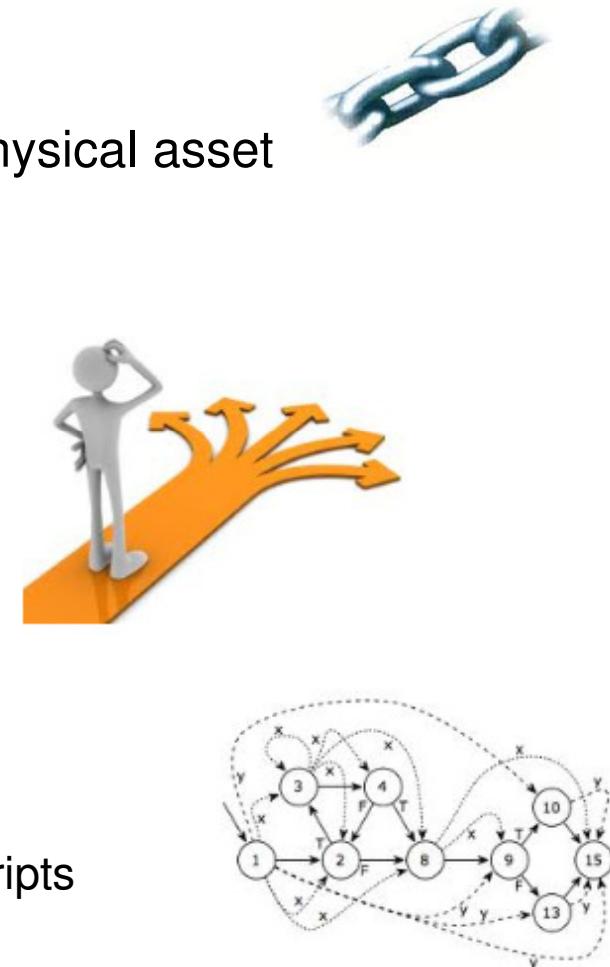
<http://www.compactbi.com/solutions/metadex>



Case Study

Business Glossary and Governance Initiative: Issues to Address

- Business Terms
 - Are not standardized
 - Cannot be defined without a corresponding physical asset
 - Cannot be linked to multiple assets
- Current Glossary
 - Is not intuitive
 - Tool and project specific
 - Not widely adopted
 - Difficult to find definitions
 - Users don't know where to go
- Lineage
 - Multiple Silos, lack of documented specs
 - Currently not trusted
 - Difficult to track
 - Teradata, INFA, Oracle, MSTR, ER Studio
 - Must be tracked through PL/SQL and BTEQ scripts



Phase I – Solution

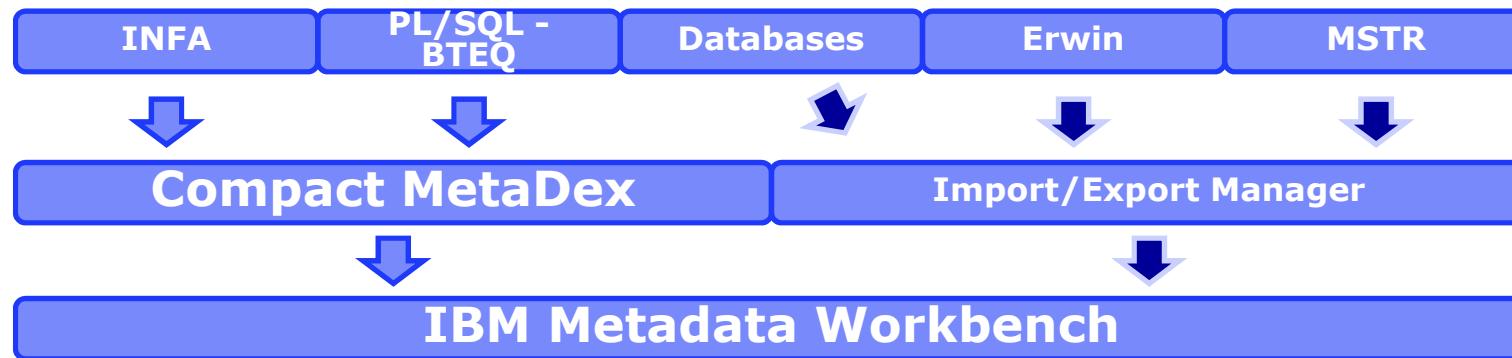
Based on requirements, the client chose:

- IBM InfoSphere Business Glossary
 - *Standardized glossary*
- Compact MetaDex
 - *Extend and automate support for additional sources*
- IBM InfoSphere Metadata Workbench
 - *Track data lineage*

InfoSphere.
software

COMPACT
SOLUTIONS

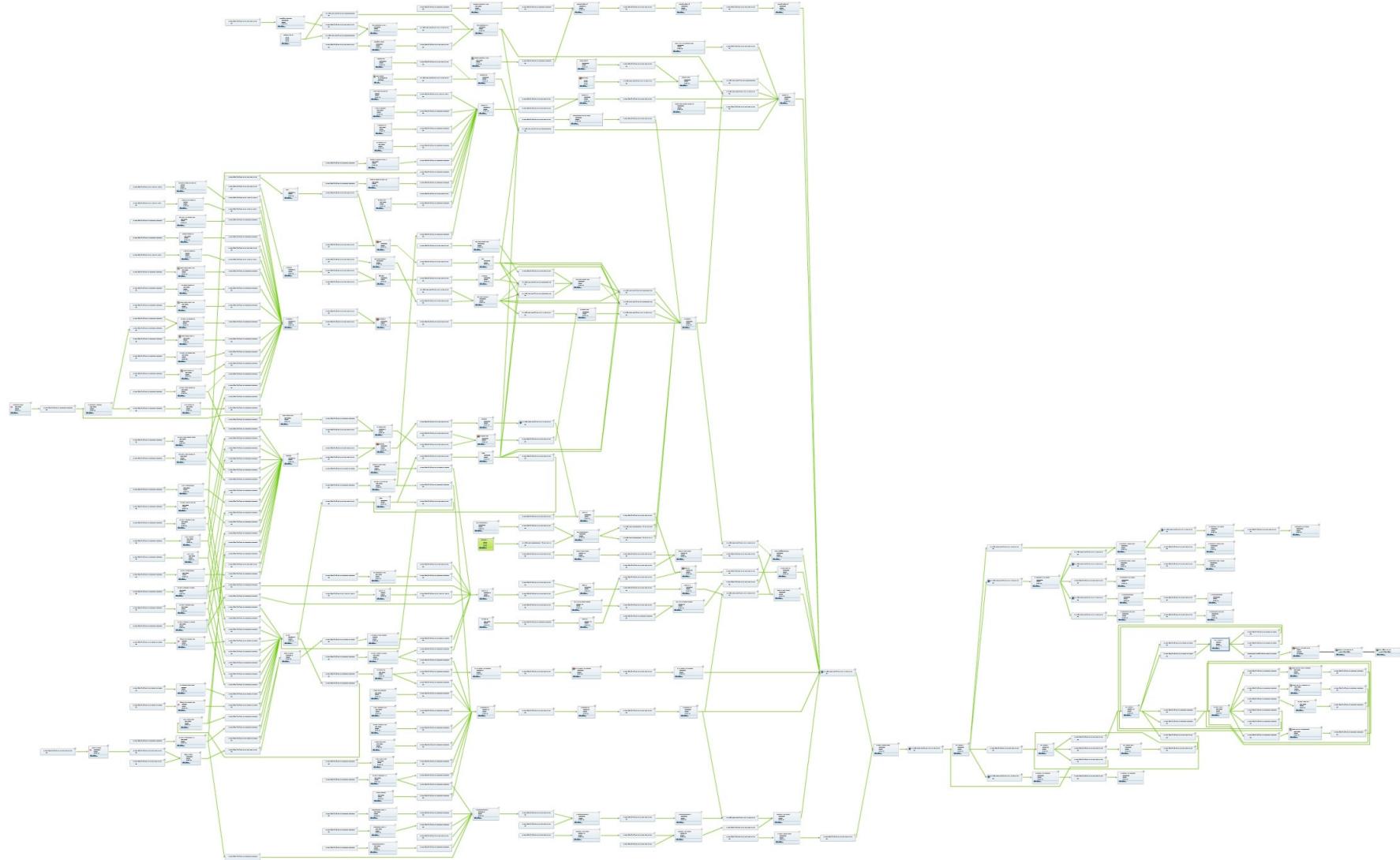
Solution Architecture



- **Deep understanding**
 - Migrate not only the physical schema, but also the processing logic itself
 - Combine knowledge about all the diverse processing in the environment, including SQL, ETL, even SQL embedded in the ETL processes
 - Take advantage of the real data lineage – field to field, expression by expression
- **Acceleration by automation**
 - Reduce time, cost and risk by automating metadata integration

How does data lineage look like?

ER Model, INFA ETL (Embedded SQL), Teradata BTEQ Scripts, MSTR Project



Business Benefits

Metadata Workbench, Business Glossary, and Compact MetaDex Solution

- **Enable better data governance**
 - Data lineage across multiple BI and ETL applications, running on different platforms
 - Rich, automated documentation of data flows and processes
 - Powerful audit capability: identify exactly where a measure is calculated in the ETL chain – which process, what transformation logic
- **Accelerate business decision making**
 - Better understanding of information at hand
 - Trust and confidence in information accuracy and consistency
 - Closer business and IT collaboration
- **Reduce costs of data integration processes**
 - Identify redundancies in processes: different processes updating same attributes or tables, unused attributes
 - Streamline ETL processes: fewer processes, better data management
 - Rapid and efficient impact analysis reduces development time

Better Business Outcomes



Respond Quickly to Compliance Demands
Use end-to-end lineage on Sarbanes-Oxley, Basel II, and other regulatory mandates.



Increase Revenue Opportunities
Make confident decisions. Understand your customers and market place.



Rescue Runaway Costs
Cut project costs by reducing development cycles, process redundancies and risks.

Thank you

