

## Information Infrastructure Foundations: Data Discovery and Mapping

***Colin Tan***  
***Regional Sales Manager, InfoSphere***



*Trusted Information*

# Data Explosion In A Smarter World

New Data For New Information For a Smarter Planet

## Volume of Digital Data

- 57% CAGR for enterprise data through 2010<sup>1</sup>
- Machine generated data : Sensors, RFID, GPS..

## Variety of Information

- 80% of new data growth is unstructured content<sup>1</sup>
- Emails, images, audio, video..

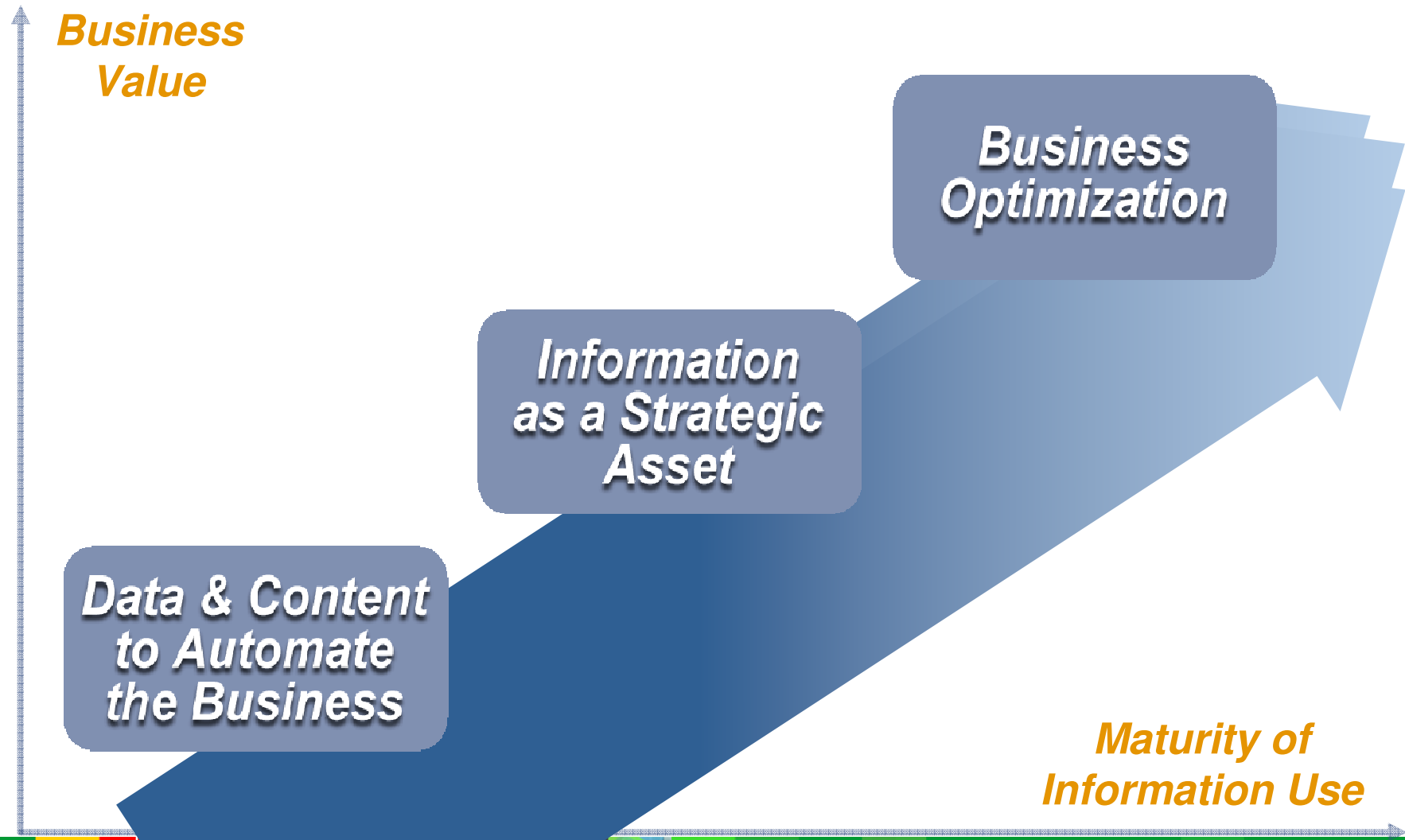
## Velocity of Decision Making

- Rapidly changing business climate
- Need to get ahead of the curve : predict issues and fix them

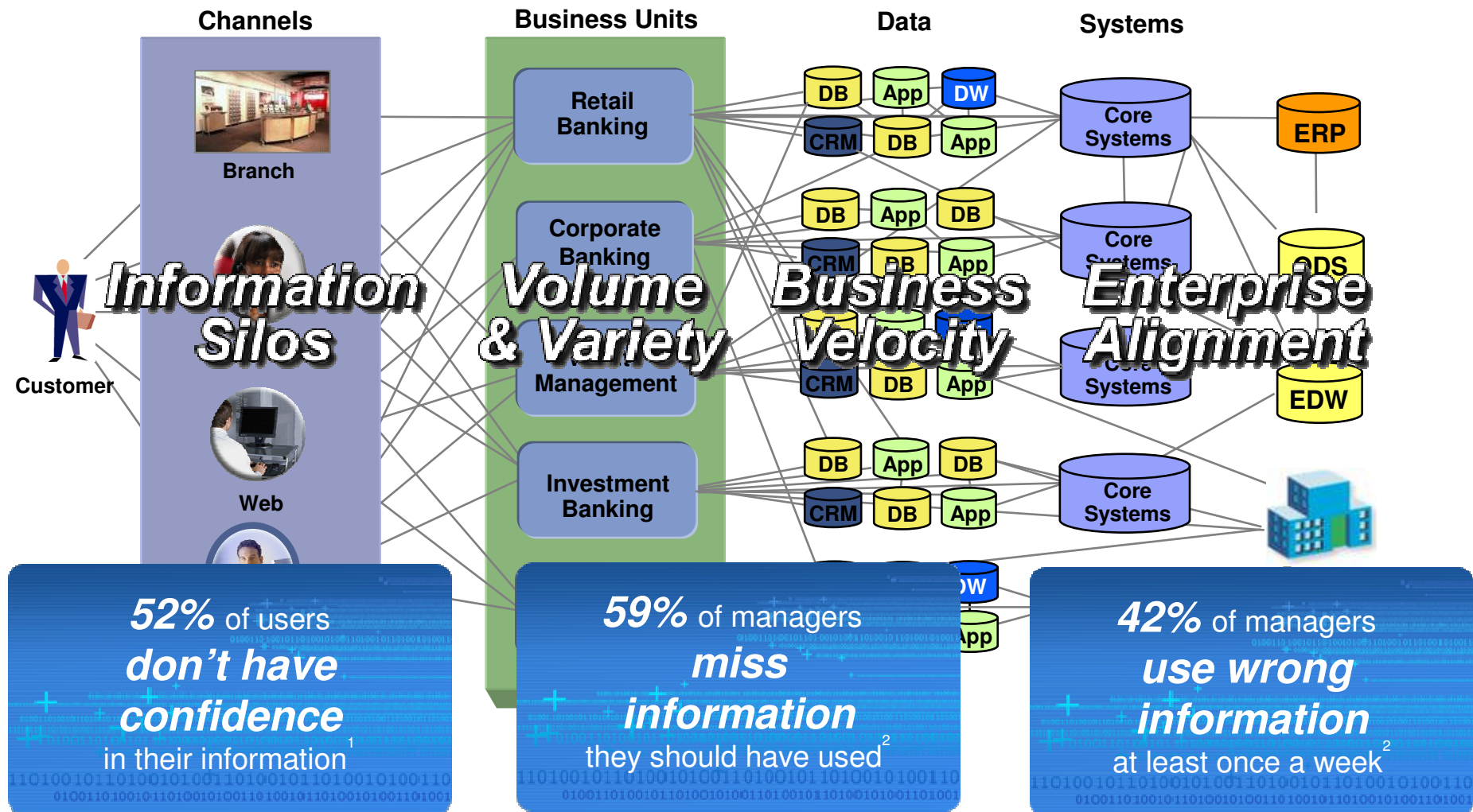


**New Data ➡ New Information!**

# Trusted Business Information Is At The Core Of Business Optimization



# Common Landscape in Banking Environments Creates an Information Gap with Strategic Challenges



<sup>1</sup>AiIM 2008 Survey

<sup>2</sup>Accenture 2007 Managers Survey  
© 2008 IBM Corporation

# Leaders Investing to Optimize Information Assets

## Clients Tell Us What They Are Looking For ...



### Insightful

Derive meaning from information changes



### In Context

Real-time delivery of relevant information when and where it's needed



### Complete

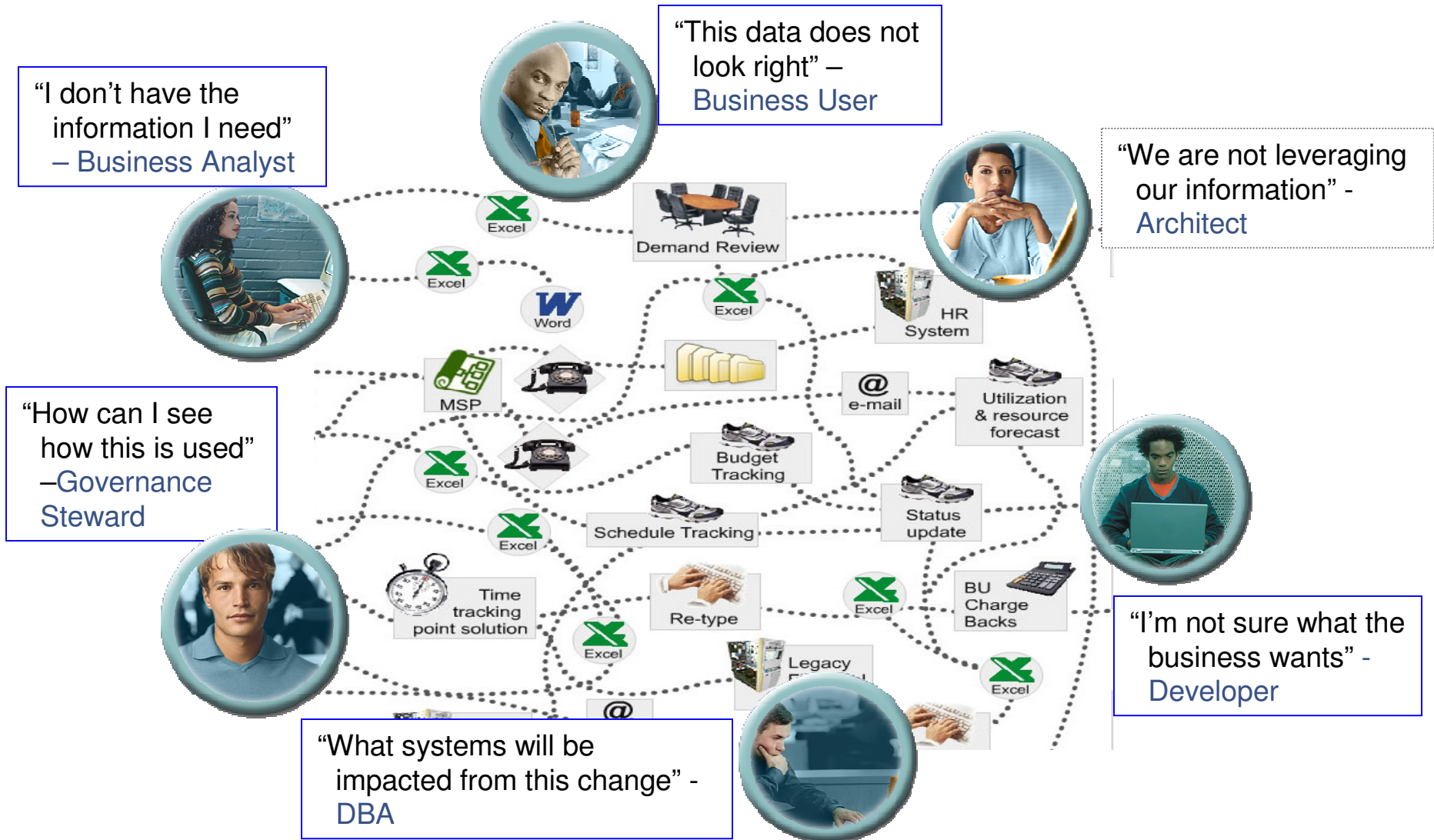
Related information reconciled into a single and holistic view



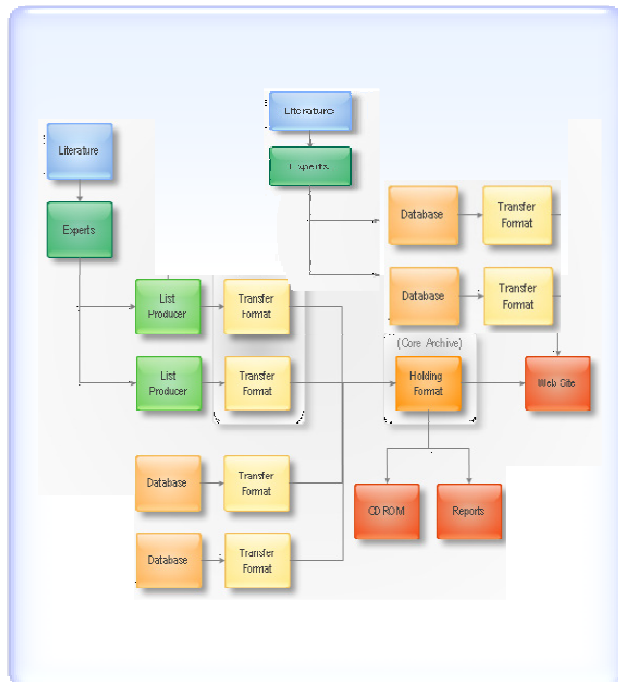
### Accurate

Complex and disparate data transformed, cleansed and delivered

# Understanding Core Information Assets Is Critical



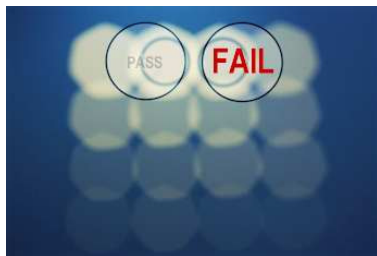
## Understanding Core Information Assets



- Which data is managed by which system?
- What relationships exist between data in different systems?
- How do I map data together for new uses?
- What should my data look like to allow me to use it to optimize my business?
- How does the business use the data?  
What meaning does it have to the business?
- Which systems are best sources for specific pieces of information?
- What data quality issues do I have?

# Impact of **NOT** Managing Core Information Assets

83% of data integration projects either overrun or fail



Scrap and rework  
Increased \$\$\$

Lack of consumer confidence

Inaccurate or incomplete data is a leading cause of failure in business-intelligence and CRM projects



Lost opportunities

Low data quality costs companies \$611 billion annually

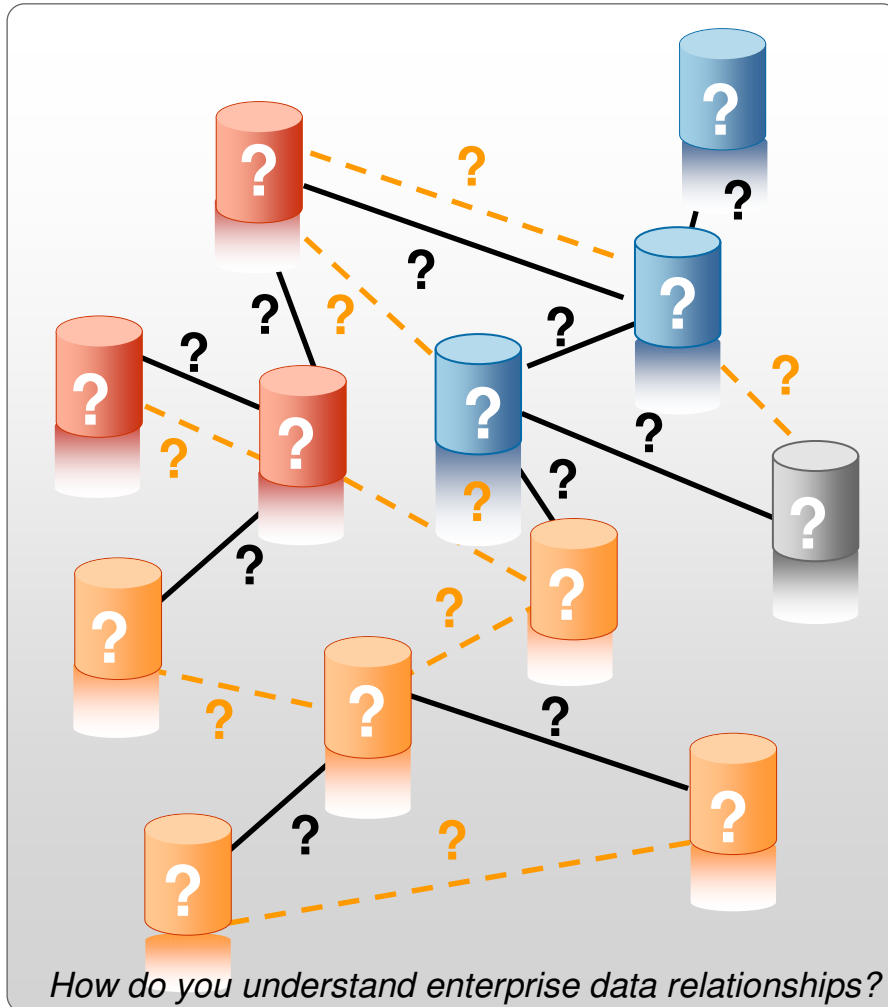
25% of time is spent clarifying bad data

Undetected defects will cost 10 to 100 times as much to fix upstream



## Challenge of understanding complex data landscapes

The first step for any successful information centric project



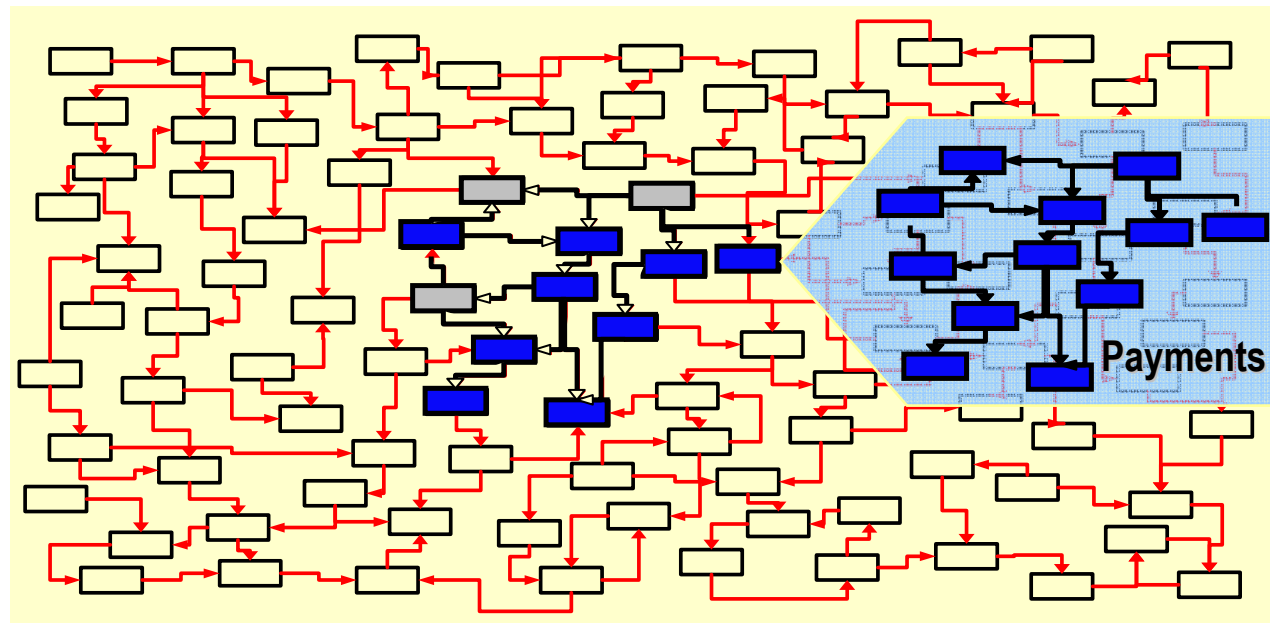
### Challenge:

- Distributed heterogeneous sources
- No documentation on data structures
- No understanding of data relationships
- Lack of trusted data – unknown quality
- Limited understanding of confidential data elements

### Cost Prohibitive Alternative Solutions:

- Manual spot checking of data
- Hand coding

## What is Data Discovery?



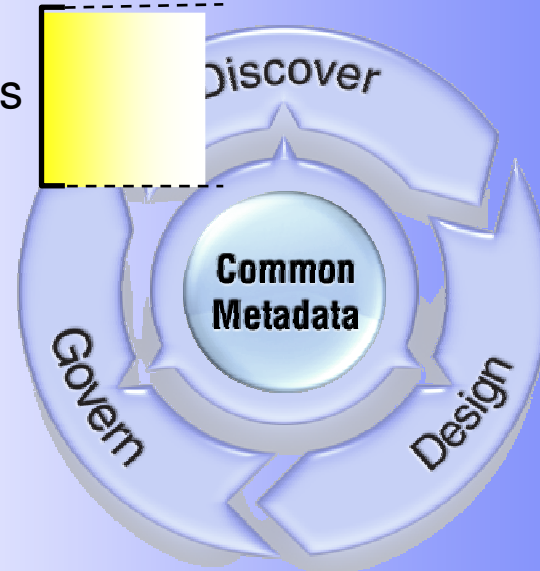
**Discovery of relationships between data elements, within and across systems – assembling a complete understanding of a business object**

*IBM acquired Exeros in May 2009 to enhance and extend the portfolio*

# Extending the Portfolio

## Enterprise Projects

**Discovery**  
Extends & Completes  
Discover Phase



InfoSphere Foundation Tools

- Test Data Generation
- Application Retirement & Consolidation
- Data Archival
- Data De-identification
- Data Quality
- Data Integration
- Master Data Management
- Data Warehousing



Manage Business Terms

**Business Glossary**



Discover Data Relationships

**New – Discovery**



Design Enterprise Models

**Data Architect**



Capture Design Specifications

**FastTrack**



Assess, Monitor, Manage Data Quality

**Information Analyzer**



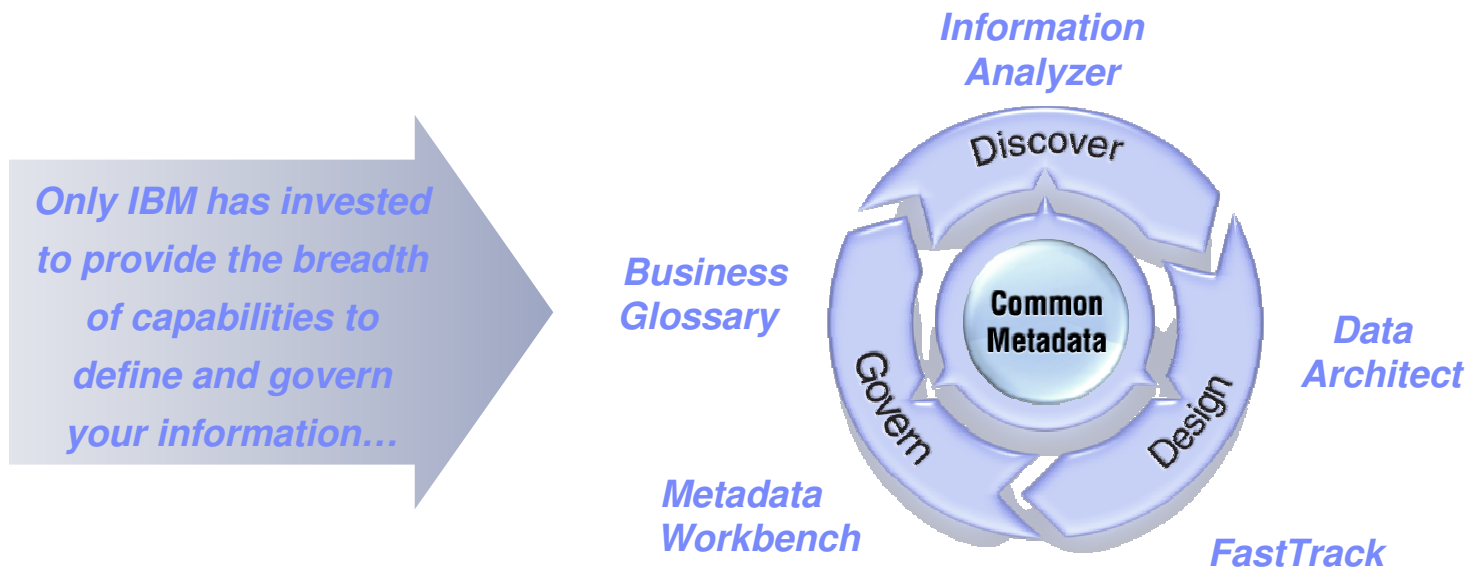
Monitor Data Flows

**Metadata Workbench**

## InfoSphere Foundation Tools

Software to help convert your information into a trusted strategic asset

*Open tools that provide value to any data integration, business intelligence, or data warehouse projects...*



- *Discover, understand and relate the data you have to your business*
- *Design your trusted information structure*
- *Govern your information over time*

# Who InfoSphere Foundation Tools are For

## Target Audience

- Data/Business Analysts
- Subject Matter Experts
- Architects
- Governance Stewards



Data/Business Analysts



Subject Matter Experts



Architects



Governance Stewards

## What are they working on?

- Information-centric projects:
  - BI & Data Warehousing
  - Master Data Management
  - Application Implementation, Consolidation or Migration
  - Information Architecture
  - Governance Initiatives

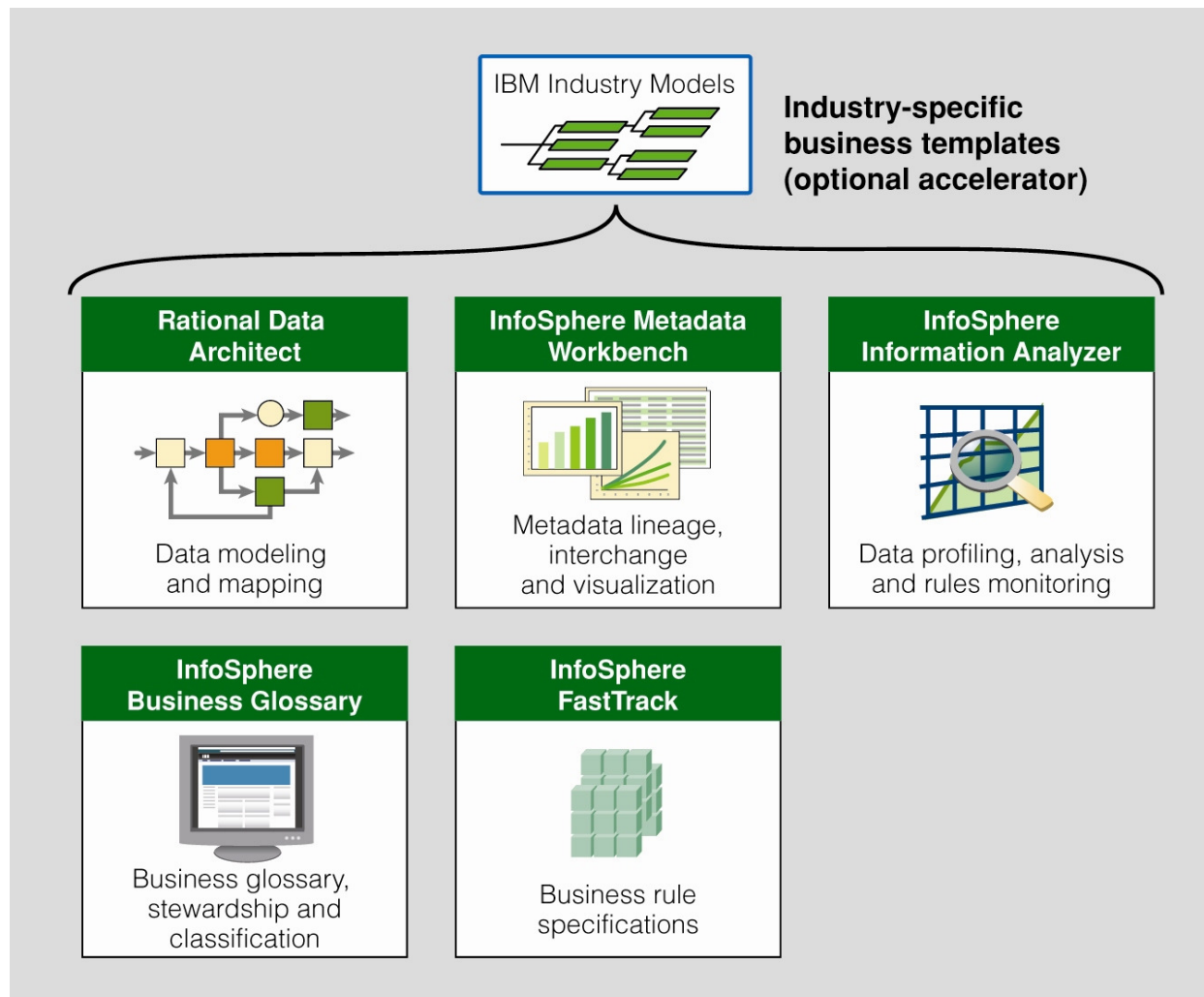
## What do these roles do today?

- Manage information manually in disconnected tools, documents, and spreadsheets

## What is wrong with what they do today?

- **Time consuming** – churn between business & IT
- **Imprecise & error prone** – manual processes not thorough enough
- **No collaboration** – different roles work in silos
- **Lacks audit trail** – no ongoing record
- **Redundancy** – duplication of effort & storage

# InfoSphere Foundation Tools Includes



# InfoSphere Foundation Tools – InfoSphere Business Glossary

- **Starting point** for designing information governance
- **Web-based authoring, sharing & management of business metadata**
- **Enhanced collaboration**
  - Aligns the goals of the business with the efforts of IT
  - Accelerate project delivery with information sharing
- **Define relationship between business definitions and IT assets**
- **Drive information trust across enterprise applications**
  - From any application, click and automatically search glossary
- **Establish accountability**
  - Creation of stewards and assignment of responsibilities



**Governance Stewards**



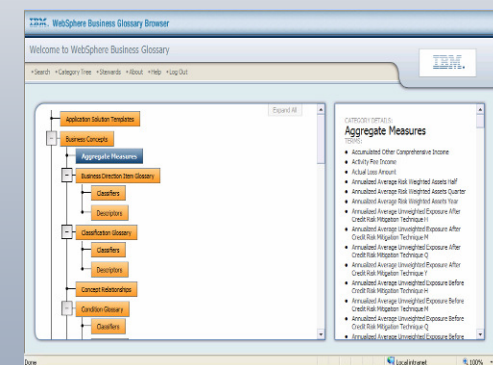
**Any User**

**Design & Govern**



**InfoSphere Business Glossary**

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



**Business Context View**

## Business Glossary Example

Verizon Partner Solutions | Glossary of Telecom Terms : M - Microsoft Internet Explorer

Address <http://www22.verizon.com/wholesale/glossary/?l=m>

**Minimum Cell Rate (MCR)**  
[ATM](#) performance parameter that specifies the minimum rate for cell transmission that a network must guarantee to a user on a given [virtual circuit](#). Also, a field in an RM cell specifying the smallest value to which the ER field can be set.

**Minutes of Use Charges**  
Usage data added to monthly bill for [UNE](#) services.

**M/L**  
See [Mobil-to-Land Traffic](#).

**Mobile Switching Center or Mobile Telecommunications Switching Office (MSC or MTSO)**  
Used by a [CMRS provider](#) for originating and terminating functions for calls to or from [end user](#) customers of the CMRS carrier. Also known as MTSO (Mobile Telecommunications Switching Office).

**Mobile-to-Land Traffic (M/L)**  
Traffic that originates on a [CMRS provider's](#) system and is terminated to a [landline](#) user.

**Modem**  
Contradiction of modulator-demodulator. A device that converts signals transmitted over communication facilities; for example, converts analog voice signals to digital signals.

**Modified Final Judgment (MEI)**



# InfoSphere Foundation Tools - InfoSphere Information Analyzer

- **First step** in defining an information infrastructure across enterprise ecosystem
- **In-depth analysis of heterogeneous information – IBM or non-IBM sources**
  - Data-centric analysis of databases, files and enterprise applications for content, quality, and structure
  - Secure, detailed profiling of fields, and relationship analysis across fields and across sources
- **On-going measurement and baseline reporting of information quality**
- **Know where information is managed across systems**
  - Capture fitness of sources and re-engineering requirements for downstream use



**Subject Matter Experts**



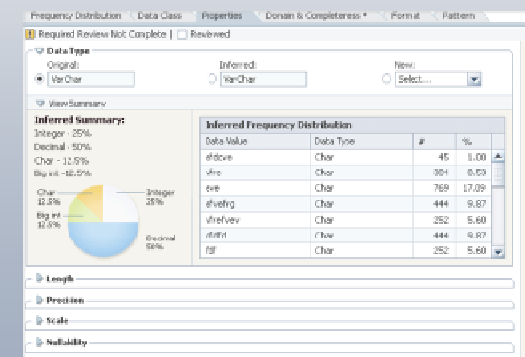
**Business & Data Analysts**

**Discover & Govern**



**InfoSphere Information Analyzer**

**Analyze source data structures, and monitor adherence to integration and quality rules**



**Physical Understanding**

# Column Analysis - Format

WORLDSCO\_BILLTO

View Analysis Summary

View Details

Select View:

- CUSTOMER\_ID
- CUSTOMER\_TYPE
- PARENT\_CUST\_ID
- PARENT\_CUST\_TYPE
- ACCT\_STATUS
- NAME
- ADDRESS\_LINE1**
- ADDRESS\_LINE2
- ADDRESS\_LINE3
- ADDRESS\_LINE4
- ADDRESS\_LINE5
- CITY
- STATE\_ABBREVIATION
- ZIP\_CODE
- COUNTRY\_CODE
- STD\_POINT\_LOC\_CODE

Frequency Distribution | Data Class | Properties | Domain & Completeness | **Format**

Number of Formats: 547

Conforming Count: 1029 | Violation Count: 0

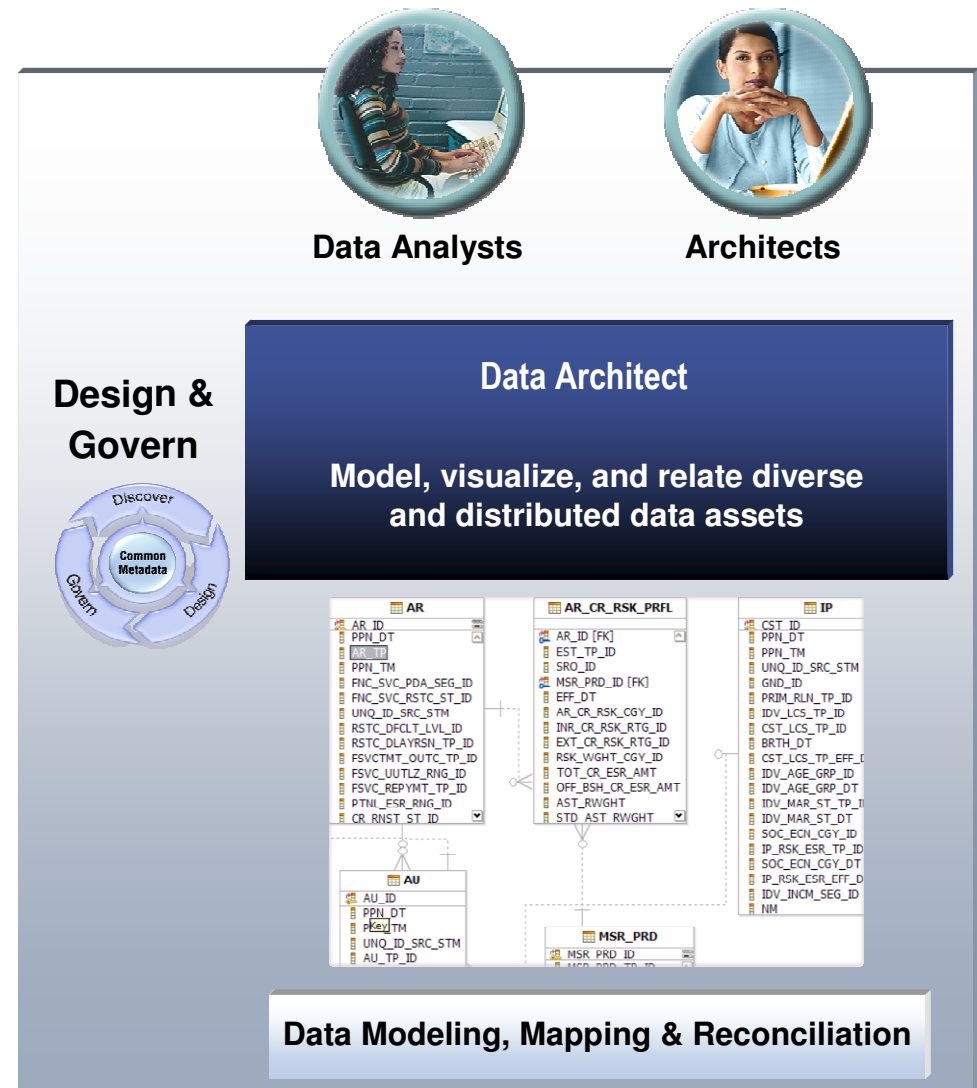
General Format				
General Format	Count	Percent	Status	
* A/A AAAAAA AAAA	29	2.82	Conform	
* A/A AAAAAAAAA AAAAAAAAA	1	0.1	Conform	
*A/A A & A AAAAAAAAAA	1	0.1	Conform	
*A/A AAAAAAAAAA AA AAAAAA	1	0.1	Conform	
.	1	0.1	Conform	
9 AAA AA	1	0.1	Conform	
9 AAAAA AAAA AAAAAA AAAA	1	0.1	Conform	
99 A 9AA AAA	1	0.1	Conform	
99 A AAAAAA AA	1	0.1	Conform	

Distinct Values		
Distinct Value	Count	Percent
* C/O MATRIX INTL	29	2.8183

Format Violations				
Distinct Value	General Format	Count	Percent	

# InfoSphere Foundation Tools - Data Architect

- **Design information infrastructure modeling assets**
  - Leverage profiling results from Information Analyzer
  - Create and optimize physical and logical data models
  - Design and deploy federated databases
- **Define enterprise standards to govern data models**
  - Analyze models for conformance
  - Compare and synchronize across models
- **Enhance collaboration**
  - Create glossary models and exchange with Business Glossary
  - SQL & XML generation capabilities



# InfoSphere Foundation Tools - InfoSphere FastTrack

- **Create a solid information infrastructure**
- **Design transformation rules to optimize business information**
  - Leverage Information Analyzer profiling results for most complete requirements
  - Flexible rules support textual descriptions or function logic
- **Centrally track design decisions for auditing**
- **Enhance collaboration**
  - Easy to use desktop interface or import work from Excel spreadsheets
  - Define and link InfoSphere Business Glossary terms to physical structures
- **Accelerate development of InfoSphere integration jobs**



**Subject Matter Experts**



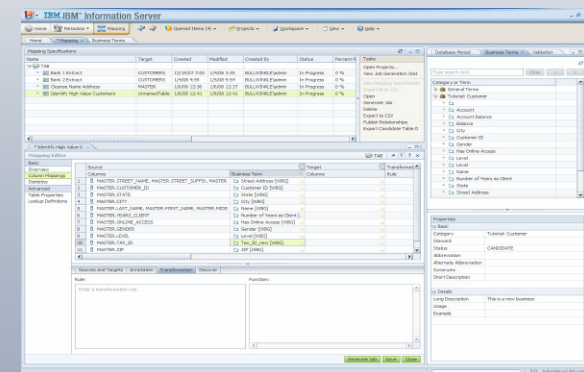
**Data & Business Analysts**

**Design**



**InfoSphere FastTrack**

**Create required transformation rules for source-to-target historical mapping documentation**



**Link Data Rules with Business Meaning**

# InfoSphere FastTrack – Sample Mappings

\* Bank 1 Extract x

Mapping Editor

	Source	Target	Business Term	Transformation
	Columns	Columns		Rule
4	CHECKING.CITY	CUSTOMERS.CITY		
5	CHECKING.CUSTOMER_ID	CUSTOMERS.CUSTOMER_ID		
6		CUSTOMERS.GENDER		setnul()
7		CUSTOMERS.LEVEL		setNull()
8	CHECKING.NAME	CUSTOMERS.NAME		
9		CUSTOMERS.ONLINE_ACCESS		setNull()
10	CHECKING.STATE	CUSTOMERS.STATE		
11	CHECKING.SS_NUM	CUSTOMERS.TAX_ID		
12		CUSTOMERS.YEARS_CLIENT		setNull()
13	CHECKING.ZIP	CUSTOMERS.ZIP		
14				


Sources and Targets | Annotation | Transformation | Discover

Rule:


Function: `setnull()`

# InfoSphere Foundation Tools – InfoSphere Metadata Workbench


- **Understand your *end-to-end information infrastructure***
- **Proactively manage and administer your information infrastructure**
  - Web-based exploration of metadata relationships – InfoSphere, modeling and BI applications
  - Explore and analyze both graphically and textually
  - Perform searches and customized queries
- **Assess and mitigate change management risk**
  - Assess dependencies across InfoSphere and 3rd party tools
- **Support compliance and governance initiatives (eg Sarbanes-Oxley, Basel II)**
  - Trace data lineage of modeling, BI reports and InfoSphere objects




**Subject Matter Experts**



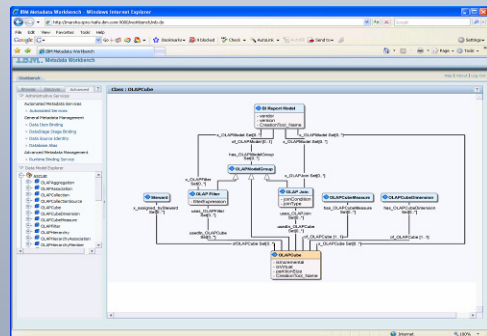
**Governance Stewards**



**Architects**



**Design & Govern**



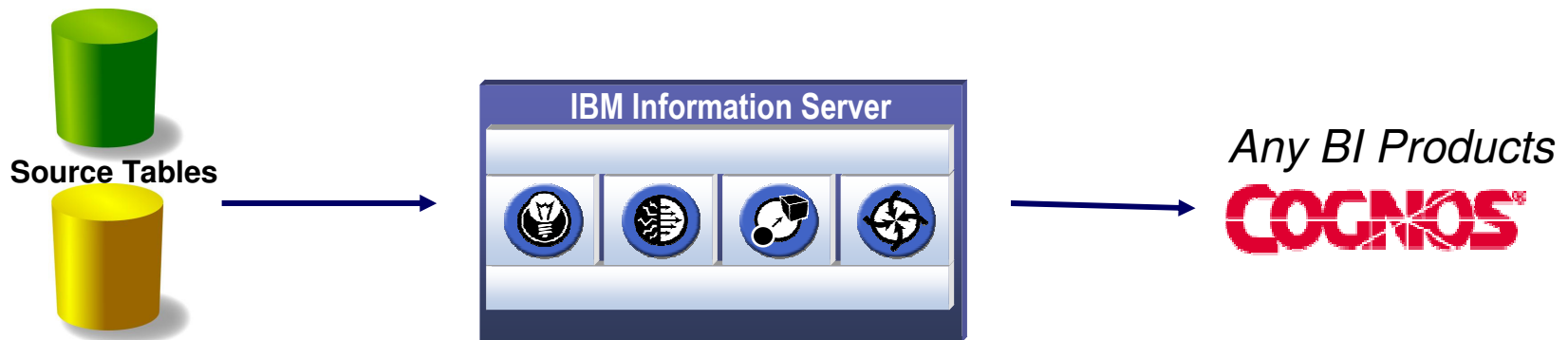
**Impact Analysis & Lineage**

**InfoSphere Metadata Workbench**

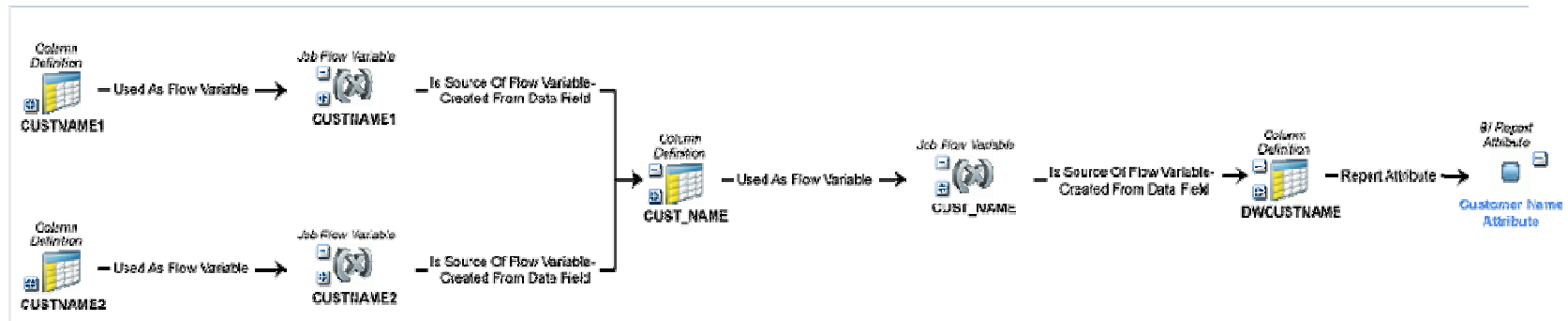
Support information governance with traceability on InfoSphere data movement, modeling & BI applications

# Where does a field of data in this report come from?

- Import & Browse Full BI Report Metadata
- Navigate through report attributes
- Visually navigate through data lineage across tools
- Combines operational & design viewpoint

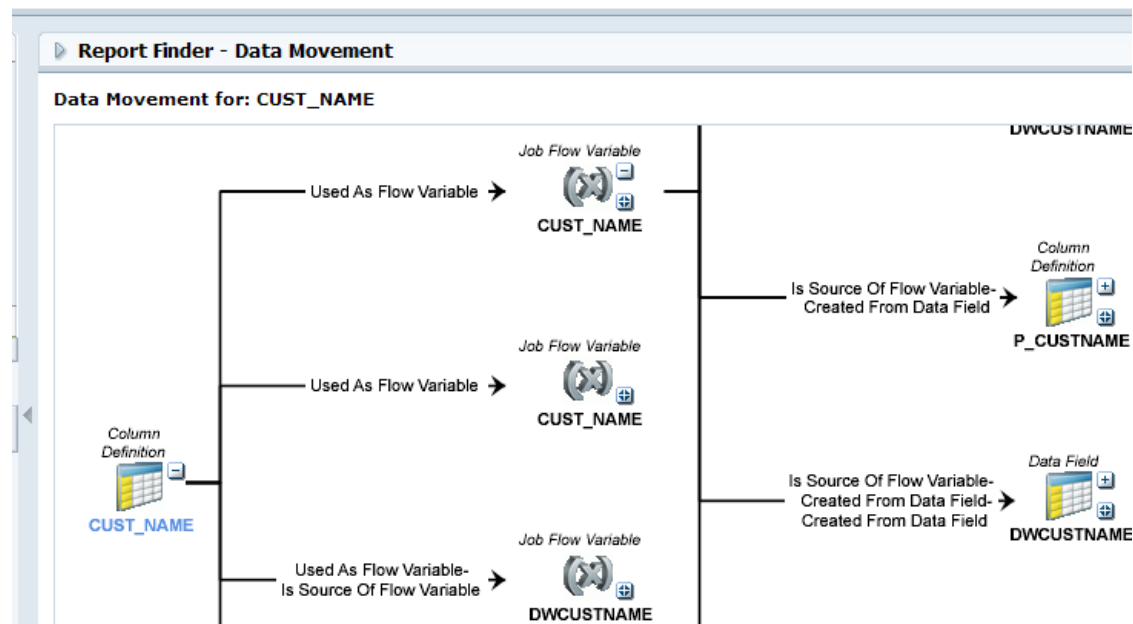


Data Movement for: Customer Name Attribute



## What happens if I change this column?

- Show **complete change** impact in graphical or list form
- **Includes impact on reports in BI tools**
- **Visually navigate through impacted objects across tools**
- **Allows impact analysis on any object type**

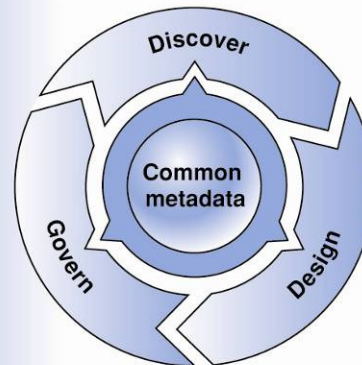




# Accelerate Foundation Tools Deployment with Industry Templates

## Accelerate deployment with industry templates

IBM Industry Models	
 <p><b>Banking</b> (Banking data warehouse)</p> <ul style="list-style-type: none"> <li>• Profitability</li> <li>• Relationship marketing</li> <li>• Risk management</li> <li>• Asset and liability management</li> <li>• Compliance</li> </ul>	 <p><b>Financial markets</b> (Financial markets data warehouse)</p> <ul style="list-style-type: none"> <li>• Risk management</li> <li>• Asset and liability management</li> <li>• Compliance</li> </ul>
 <p><b>Health plan</b> (Health plan data warehouse)</p> <ul style="list-style-type: none"> <li>• Claims</li> <li>• Medical management</li> <li>• Provider and network</li> <li>• Sales, marketing and membership</li> <li>• Financials</li> </ul>	 <p><b>Insurance</b> (Insurance information warehouse)</p> <ul style="list-style-type: none"> <li>• Customer centricity</li> <li>• Claims</li> <li>• Intermediary performance</li> <li>• Compliance</li> <li>• Risk management</li> </ul>
 <p><b>Retail</b> (Retail data warehouse)</p> <ul style="list-style-type: none"> <li>• Customer centricity</li> <li>• Merchandising management</li> <li>• Store operations and product management</li> <li>• Supply chain management</li> <li>• Compliance</li> </ul>	 <p><b>Telco</b> (Telecommunications data warehouse)</p> <ul style="list-style-type: none"> <li>• Churn management</li> <li>• Relationship management and segmentation</li> <li>• Sales and marketing</li> <li>• Service quality and product lifecycle</li> <li>• Usage profile</li> </ul>



- Provides massive **acceleration** for your **Information Agenda**
- Pre-defined proven industry models across **six industries**
- Robust data models form a complete **foundation for design**
- **Pre-populate Business Glossary** with full industry vocabulary

# Banking Data Warehouse Model

## Supported Analysis using Business Solution Template

	<b>Customer Relationship</b>	<ul style="list-style-type: none"> <li>• Customer Interaction Analysis</li> <li>• Customer Investment Profile</li> <li>• Individual Customer Profile</li> <li>• Wallet Share Analysis</li> </ul>	<ul style="list-style-type: none"> <li>• Customer Complaints</li> <li>• Delinquency Analysis</li> <li>• Customer Loyalty</li> <li>• Market Analysis</li> </ul>	<ul style="list-style-type: none"> <li>• Campaign Analysis</li> <li>• Cross Sell Analysis</li> <li>• Customer Attrition</li> <li>• Lead Analysis</li> <li>• Customer Behavior</li> </ul>
	<b>Profitability</b>	<ul style="list-style-type: none"> <li>• Transaction Analysis</li> <li>• Activity Based Costing Analysis</li> <li>• Insurance Product Analysis</li> <li>• Investment Arrangement Analysis</li> </ul>	<ul style="list-style-type: none"> <li>• Profitability Analysis</li> <li>• Channel Profitability</li> <li>• Customer Lifetime Value</li> <li>• Customer Profitability</li> </ul>	<ul style="list-style-type: none"> <li>• Product Profitability</li> <li>• Product Analysis</li> <li>• Organization Unit Profitability</li> <li>• Performance Measurement</li> </ul>
	<b>Risk</b>	<ul style="list-style-type: none"> <li>• Credit Risk Profile</li> <li>• Credit Risk Assessment</li> <li>• Asset Securitization Analysis</li> <li>• Operational Risk Assessment</li> </ul>	<ul style="list-style-type: none"> <li>• Outstandings Analysis</li> <li>• Portfolio Credit Exposure</li> <li>• Security Analysis</li> <li>• Liquidity Risk</li> </ul>	<ul style="list-style-type: none"> <li>• Debt Restructuring</li> <li>• Involved Party Exposure</li> <li>• Location Exposure</li> <li>• Non Performing Loan</li> </ul>
	<b>Asset &amp; Liability Management</b>	<ul style="list-style-type: none"> <li>• Interest Rate Sensitivity</li> <li>• Liquidity Analysis</li> <li>• Short Term Funding Management</li> <li>• Financial Management Accounting</li> </ul>	<ul style="list-style-type: none"> <li>• Capital Allocation Analysis</li> <li>• Capital Procurement</li> <li>• Credit Loss Provision</li> <li>• Funds Maturity Analysis</li> </ul>	<ul style="list-style-type: none"> <li>• Net Interest Margin Variance</li> <li>• Structured Finance Analysis</li> <li>• Equity Position Exposure</li> <li>• Income Analysis</li> </ul>
	<b>Compliance</b>	<ul style="list-style-type: none"> <li>• Financial Capital Adequacy Analysis</li> <li>• Structure Of Regulatory Capital</li> <li>• Suspicious Activity Analysis</li> <li>• SOX Balance Sheet Analysis</li> </ul>	<ul style="list-style-type: none"> <li>• Cash Flow Indirect Financial Institution Analysis</li> <li>• Income Statement By Function Analysis</li> <li>• Statement Of Changes In Equity Analysis</li> <li>• Balance Sheet Portfolio Basis Approach Analysis</li> </ul>	

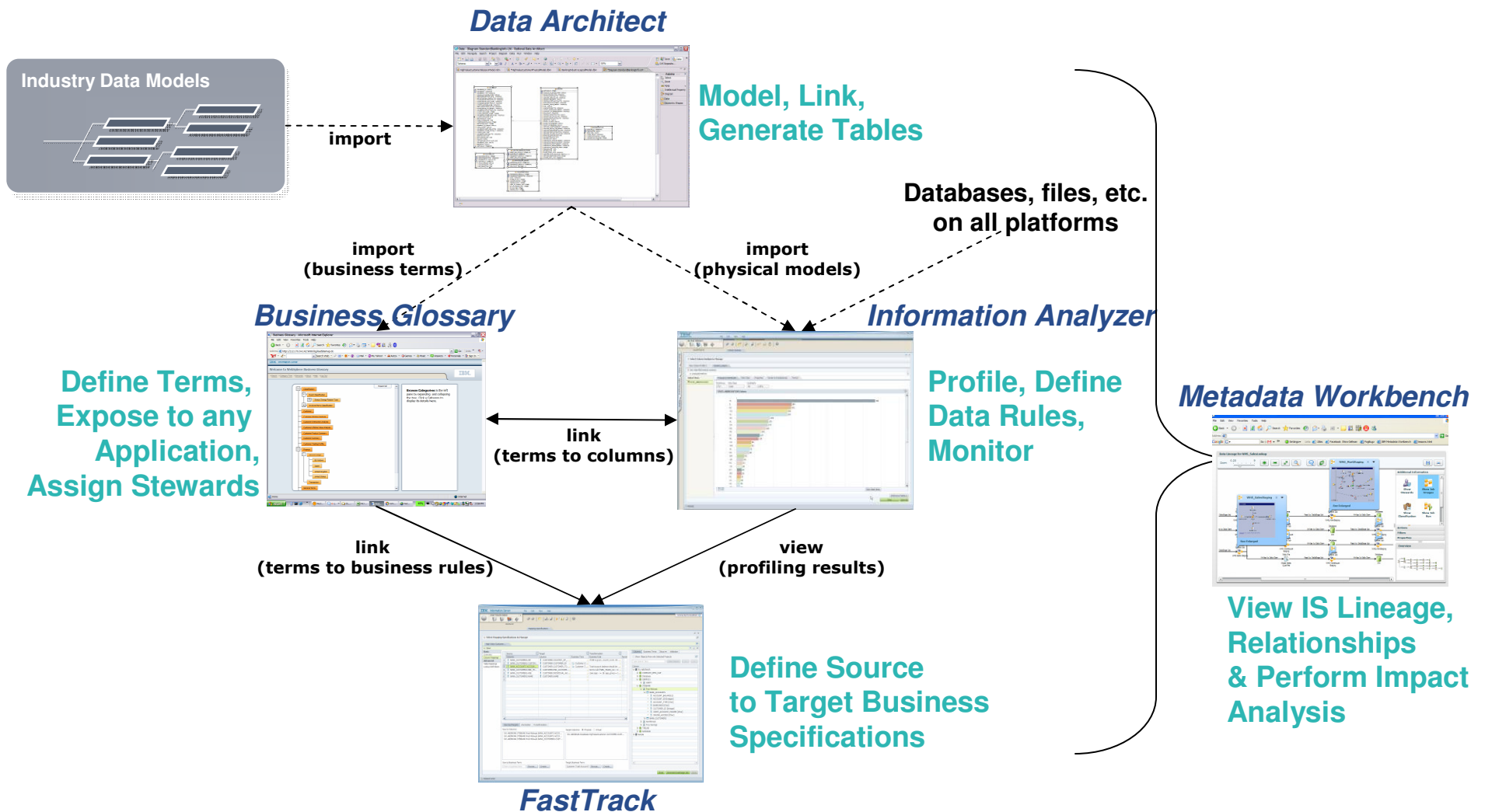
Please refer to IFW BDW GIM for a complete BST listing

# Telecommunications Data Warehouse Model

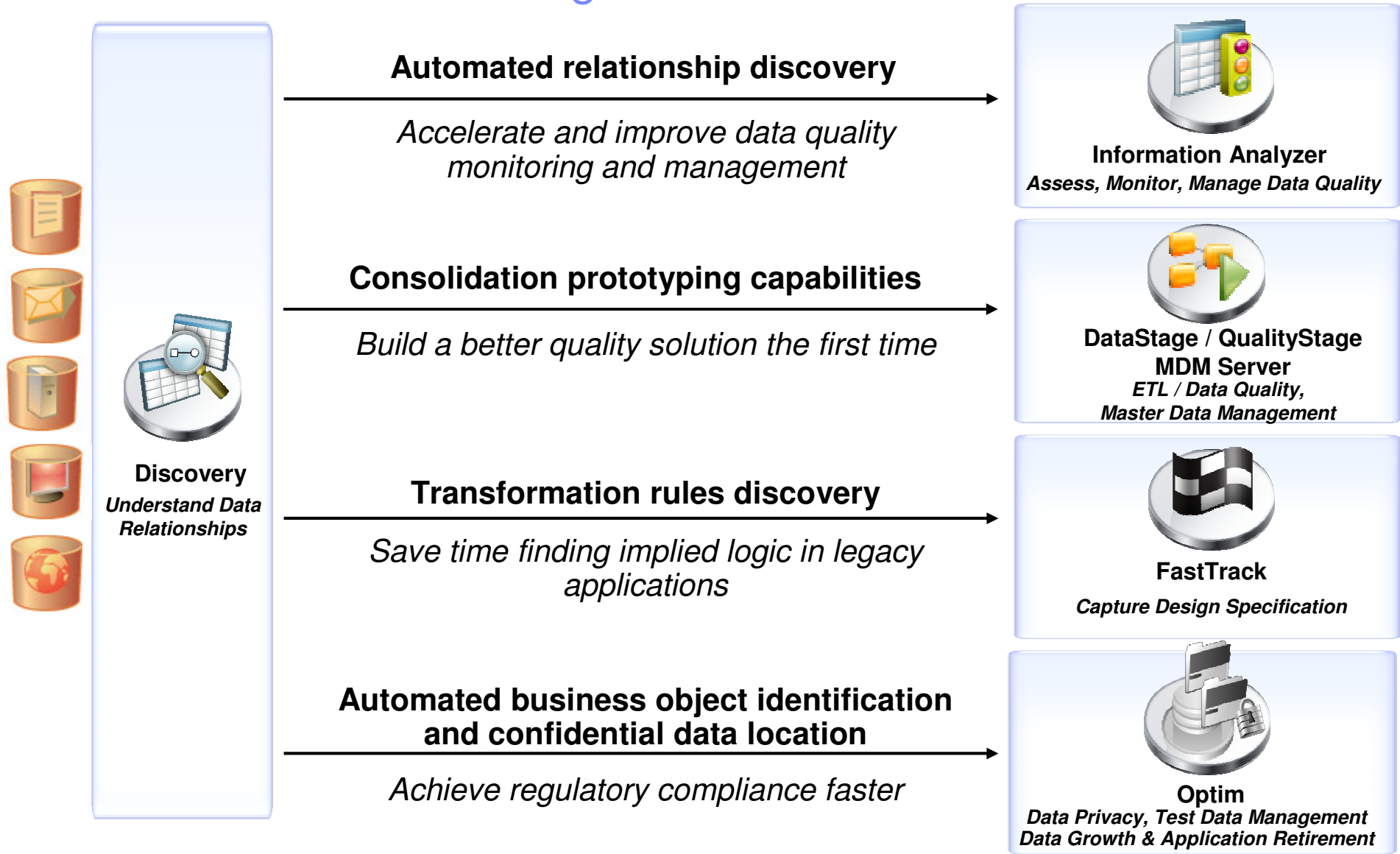
## Supported Analysis using Business Solution Templates (TBST)

	<b>Sales and Marketing</b>	<ul style="list-style-type: none"> <li>• Campaign Analysis</li> <li>• Cross Sell Analysis</li> <li>• Customer Acquisition Analysis</li> <li>• Data Package Sales Analysis</li> </ul>	<ul style="list-style-type: none"> <li>• Number Profitability Analysis</li> <li>• Retail Transaction Analysis</li> <li>• Sales Channel Analysis</li> <li>• MSO Advertising Inventory &amp; Sales Mgmt</li> </ul>
	<b>Usage Profiling</b>	<ul style="list-style-type: none"> <li>• Content Usage Analysis</li> <li>• E-Commerce Analysis</li> <li>• In/Outbound Roamer Usage Analysis</li> <li>• Pre-rate CDR Analysis</li> </ul>	<ul style="list-style-type: none"> <li>• SMS/MMS/PPV/VoD Usage</li> <li>• Wireless Data/Voice Usage</li> <li>• Wireline Data/Voice Usage</li> <li>• MVNO Data/Voice Usage</li> <li>• Stock Availability Analysis</li> </ul>
	<b>Relationship Management/Segmentation</b>	<ul style="list-style-type: none"> <li>• Commercial Customer Analysis</li> <li>• Customer Arrangement Analysis</li> <li>• Customer Lifetime Value Analysis</li> <li>• Customer Profitability Analysis</li> </ul>	<ul style="list-style-type: none"> <li>• Individual Customer Financial Analysis</li> <li>• Customer Complaints Analysis</li> <li>• Wallet Share Analysis</li> <li>• Fraud Analysis</li> </ul>
	<b>Network, Service Quality And Product Lifecycle</b>	<ul style="list-style-type: none"> <li>• Customer Billing Analysis</li> <li>• Financial Management Analysis</li> <li>• Income Analysis</li> <li>• Complaint Financial Analysis</li> <li>• MRO Operations Analysis</li> </ul>	<ul style="list-style-type: none"> <li>• Product Profitability Analysis</li> <li>• Postpaid Revenue Analysis</li> <li>• Prepaid Revenue Analysis</li> <li>• Outlet Location Profitability A.</li> <li>• New Product Introduction Analysis</li> <li>• Cash Flow Analysis</li> <li>• Balance Sheet Analysis</li> <li>• Shareholder Equity A.</li> <li>• MVNO SLA compliance</li> <li>• MVNO Financial Analysis</li> </ul>
	<b>Churn Management</b>	<ul style="list-style-type: none"> <li>• Contract Renewal Analysis</li> <li>• Customer Churn Analysis</li> </ul>	<ul style="list-style-type: none"> <li>• NLD/IDD Defection Analysis</li> </ul>

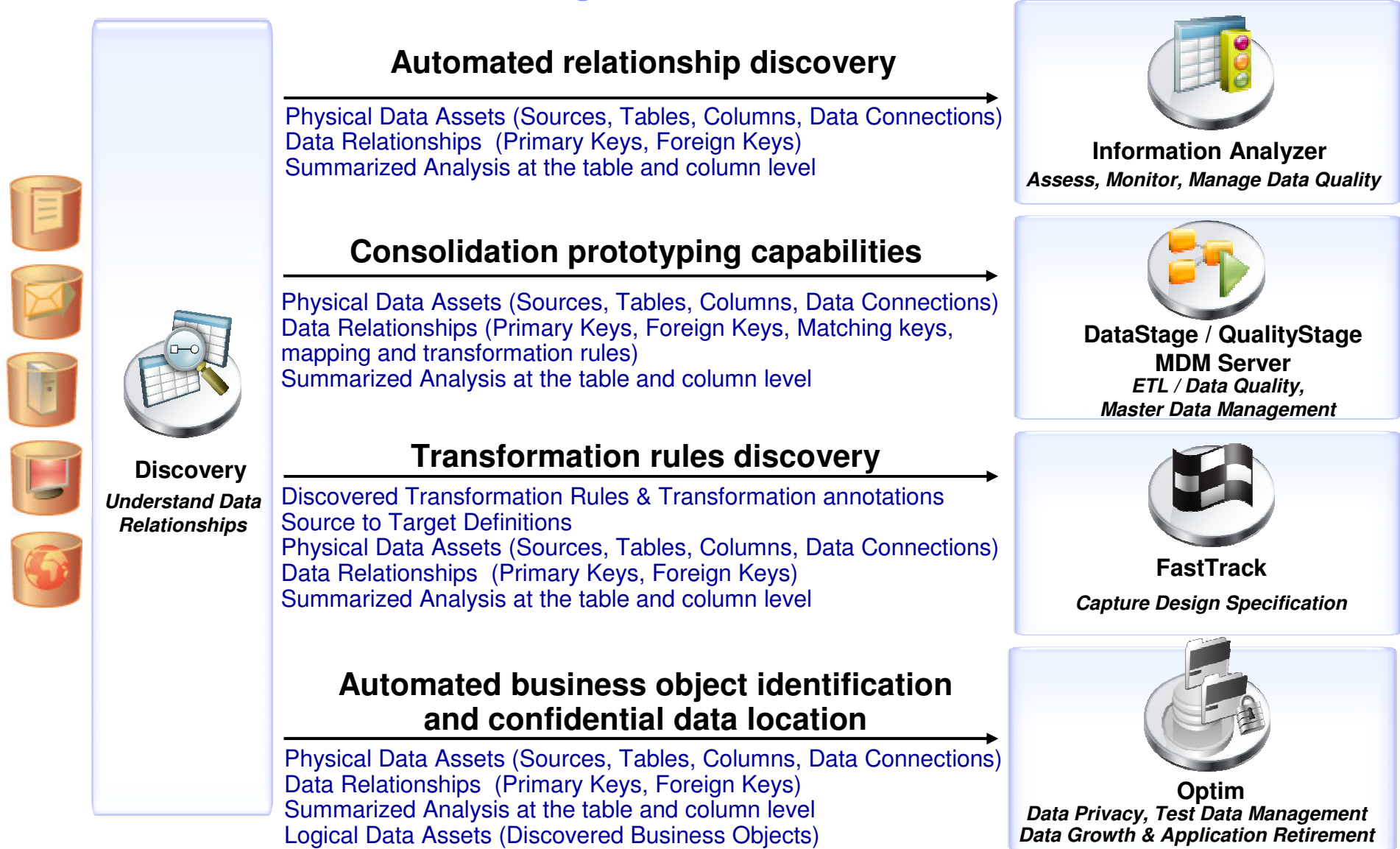
# InfoSphere Foundation Tools Integration



# How the Products Work Together



# How the Products Work Together – Metadata Integration



THANK  
YOU