



**IBM**

Information



Demand **2010**

# Intelligent Analytics and Optimization for Smarter Business



**LEAD  
THE WAY**



# Deliver Trusted Information Whenever Needed

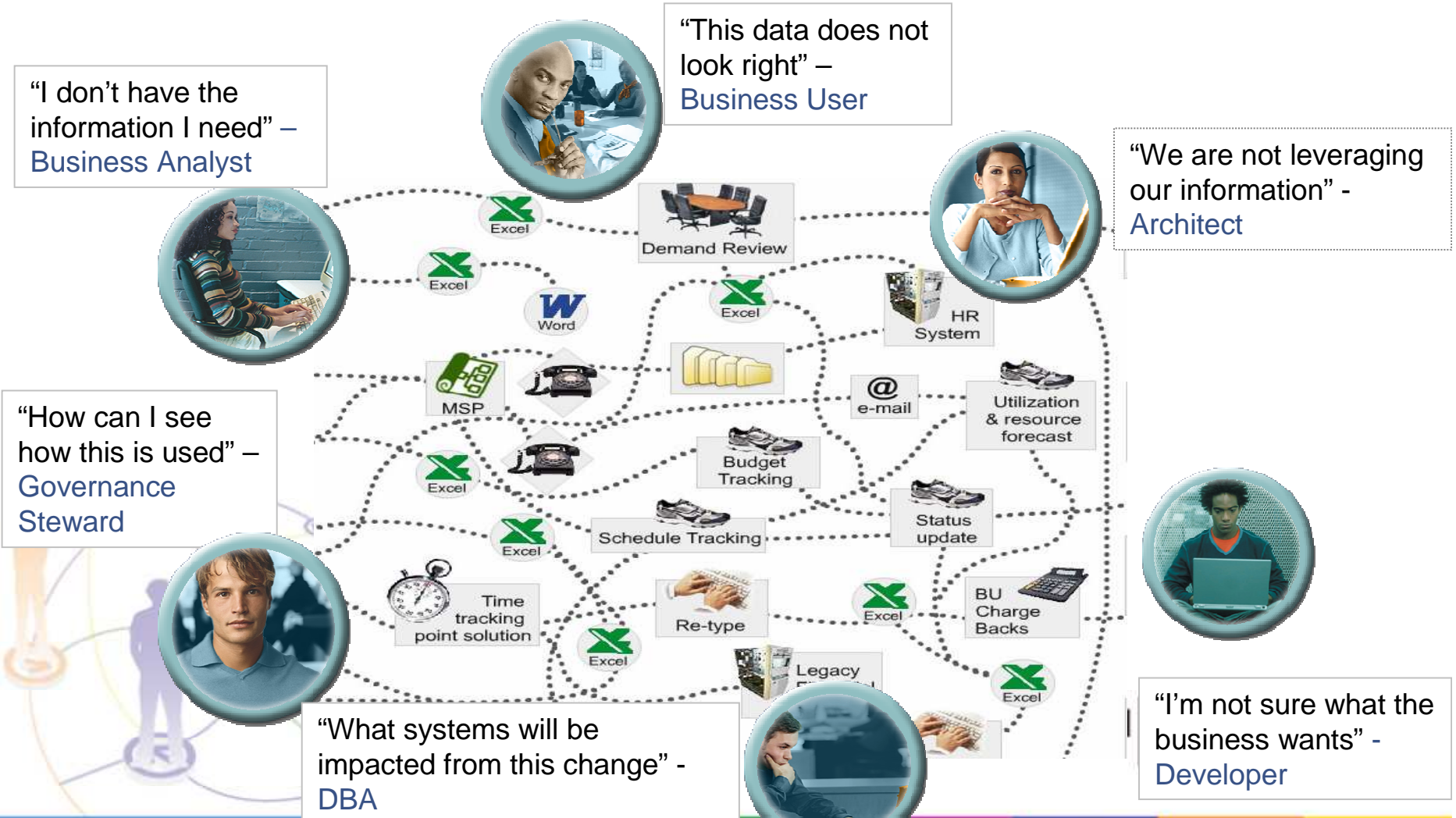
Andrew TC Lau

*Senior Sales Specialist, InfoSphere, IBM*



IBM Information ON Demand 2010

# Pain Point: Understanding Core Information Assets



# Who are Looking for Trusted Information?

## Target Audience

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



Data/Business Analysts



Subject Matter Experts



Architects



Governance Stewards

## What do these roles do today?

### Trusted Information

1. Accurate
2. Complete
3. Insightful
4. Real Time

## What

- Information Architecture
- Governance Initiatives

- **Lacks audit trail** – no ongoing record
- **Redundancy** – duplication of effort & storage

IBM Information Demand 2010

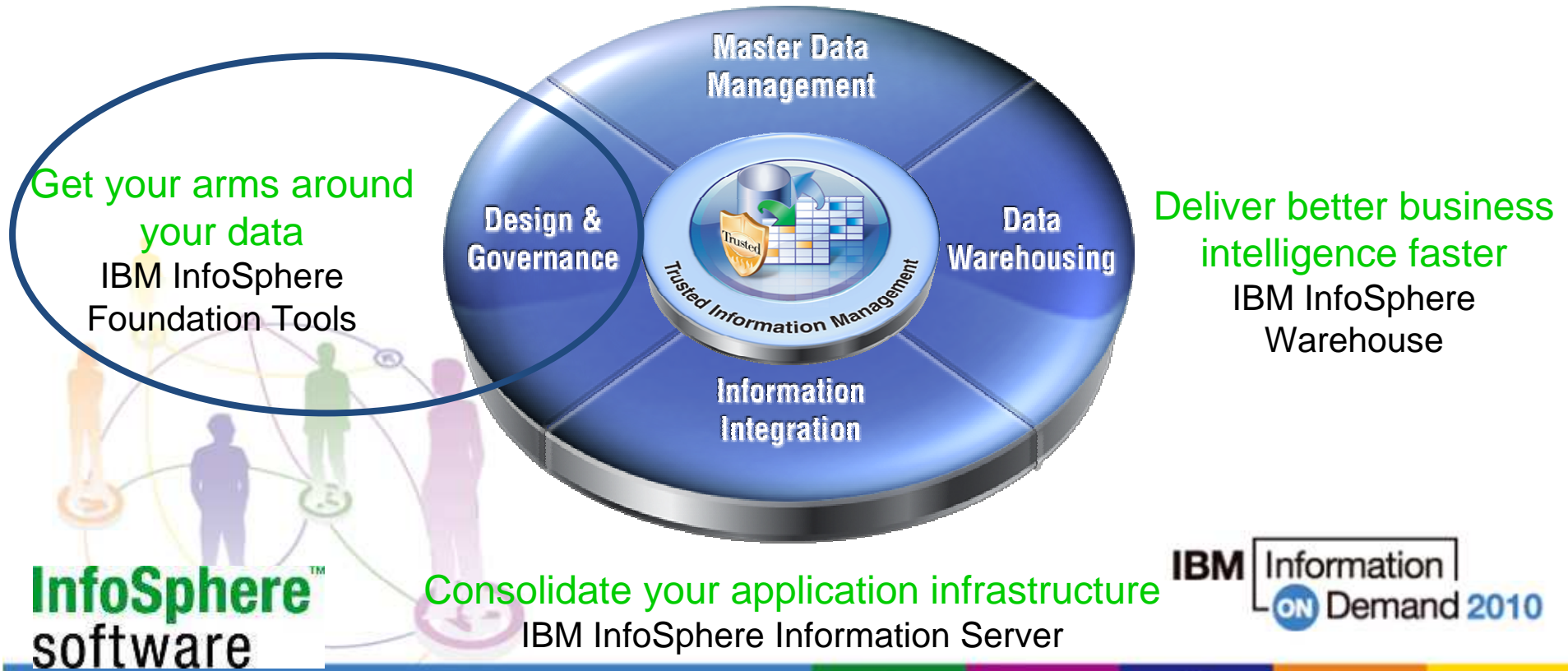




# Trusted Information Management

*Create, manage, govern and deliver trusted information*

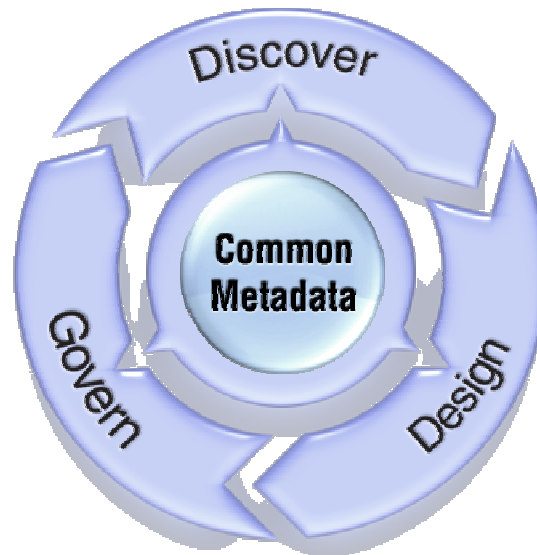
Get A single view of your business  
IBM InfoSphere MDM Server



# InfoSphere Foundation Tools

*Convert information into a trusted strategic asset*

Only IBM has invested to provide the breadth of capabilities to define and govern your information...

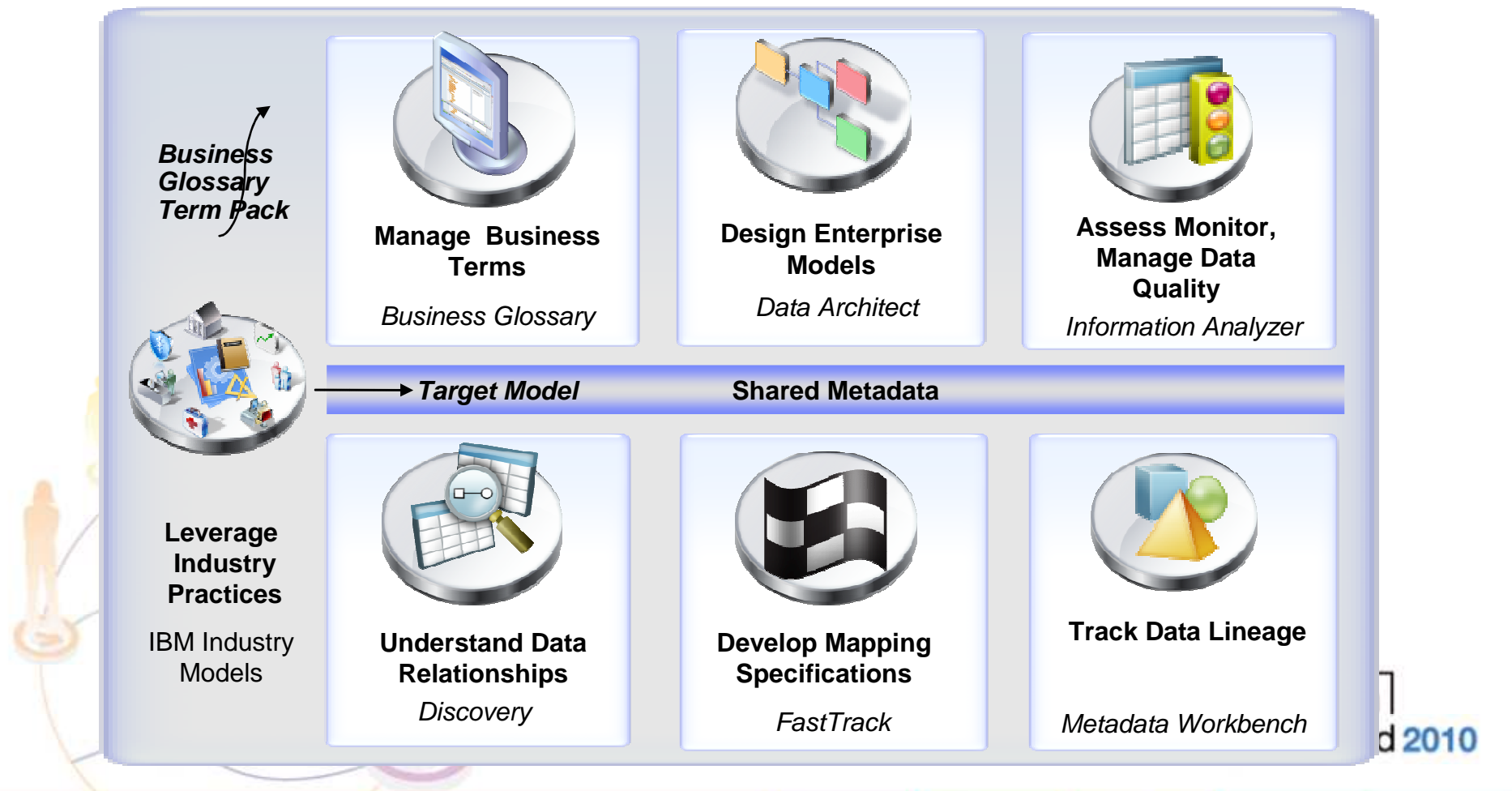


- Business Vocabulary
- Data Relationships
- Data Quality Compliance
- Data Models and Mapping
- Business Specification Rules
- Provenance of information

- **Discover** and understand the data across heterogeneous systems
- **Design** trusted information structures for business optimization
- **Govern** that information over time

IBM Information ON Demand 2010

# InfoSphere Foundation Tools



© 2010

# InfoSphere Foundation Tools

## Enterprise Projects



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

<b>Manage Business Terms</b>	<b>Discover Data Relationships</b>	<b>Design Enterprise Models</b>	<b>Capture Design Specifications</b>	<b>Assess, Monitor, Manage Data Quality</b>	<b>Monitor Data Flows</b>
<i>Business Glossary</i>	<i>Discovery</i>	<i>Data Architect</i>	<i>FastTrack</i>	<i>Information Analyzer</i>	<i>Metadata Workbench</i>





# 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



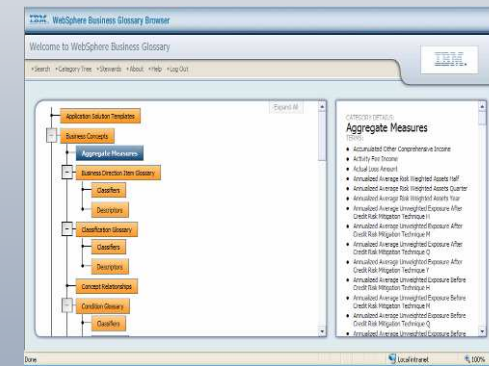
Any User

## Design & Govern



## InfoSphere Business Glossary

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



Business Context View

# Create a Common Vocabulary

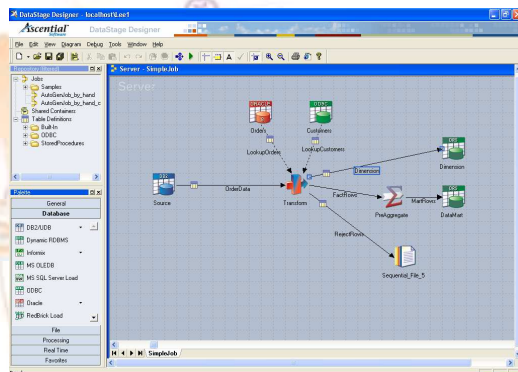
**Database = DB2**  
**Schema = NAACCT**  
**Table = DLYTRANS**  
**Column = TAXVL**  
**data type = Decimal**  
**(14,2)**  
**Derivation: SUM(TRNTXAMT)**



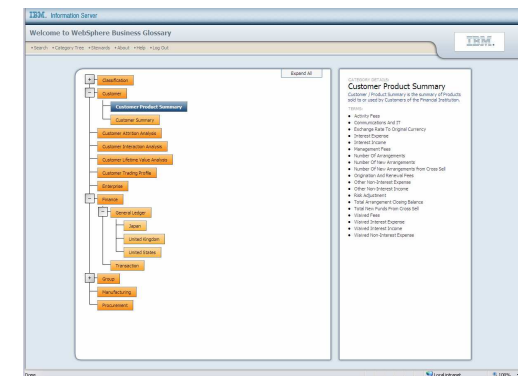
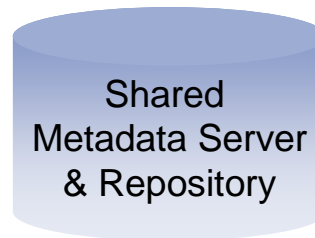
**Category: Costs**  
**Term: Tax Expense**  
**Full Name: Tax to be paid on Gross Income**  
**“The expense due to taxes .....”**  
**(John Walsh is responsible for updates. 90% reliable source)**  
**Status: CURRENT**



**Achieve a common vocabulary between business & technical users!**



InfoSphere DataStage



InfoSphere Business Glossary

# InfoSphere Business Glossary Anywhere

**1. Highlight term Active and click to look up in Business Glossary**

**2. Candidate terms returned.**

**3. Select term to view full details.**

FirstName	LastName	
Andrew	Fuller	Vice Sale
Anne	Dodsworth	Sale Rep
Janet	Leverling	Sale Rep
Laura	Callahan	Insur Cod
Margaret	Peacock	Sale Rep
Michael	Suyama	Sale Rep
Nancy	Davolio	Sale Rep
Robert	King	Sale Rep
Steven	Buchanan	Sale

**"active" - 1 Result**

**Active**  
Active refers to an employee's insurance coverage status. It does not imply their employment status.

If your term is not listed, also [search term descriptions](#).

**Employee » Active**

**Active**

Describes an employee who is currently receiving insurance coverage.

**Steward:** Richard Smith  
**Status:** STANDARD

**General Information**

**Long Description**  
↳ Active refers to an employee's insurance coverage status. It does not imply their employment status.

**Usage**  
↳ Active employees each receives monthly insurance newsletters.



# Cognos Integration with Glossary

The screenshot shows the Cognos Query Studio interface within a Microsoft Internet Explorer browser window. The browser address bar shows `http://localhost/cognos8/cgi-bin/cognos.cgi`. The main interface includes a menu on the left with options like 'Insert Data', 'Edit Data', 'Change Layout', 'Run Report', and 'Manage File'. Below the menu is a tree view of data sources, including 'GO Data Warehouse (analysis)' and 'HR (analysis)'. A 'Term Details' window is open, displaying information for the term 'High Value Customer'. The details include a description: 'Any customer with a combined account balance exceeding \$500,000', a steward 'Data Steward Roger Hecker', and a status of 'STANDARD'. The 'General Information' section lists 'Long Description', 'Synonyms', 'Preferred Synonym', 'Abbreviation', 'Additional Abbreviation', 'Categories that Refer to this Term', 'Example', 'Related Terms', 'Replaced By Term', 'Replaces Term', and 'Usage'. A red circle highlights the 'Glossary' option in the 'Insert' menu, with a blue arrow pointing from it to the 'Term Details' window.



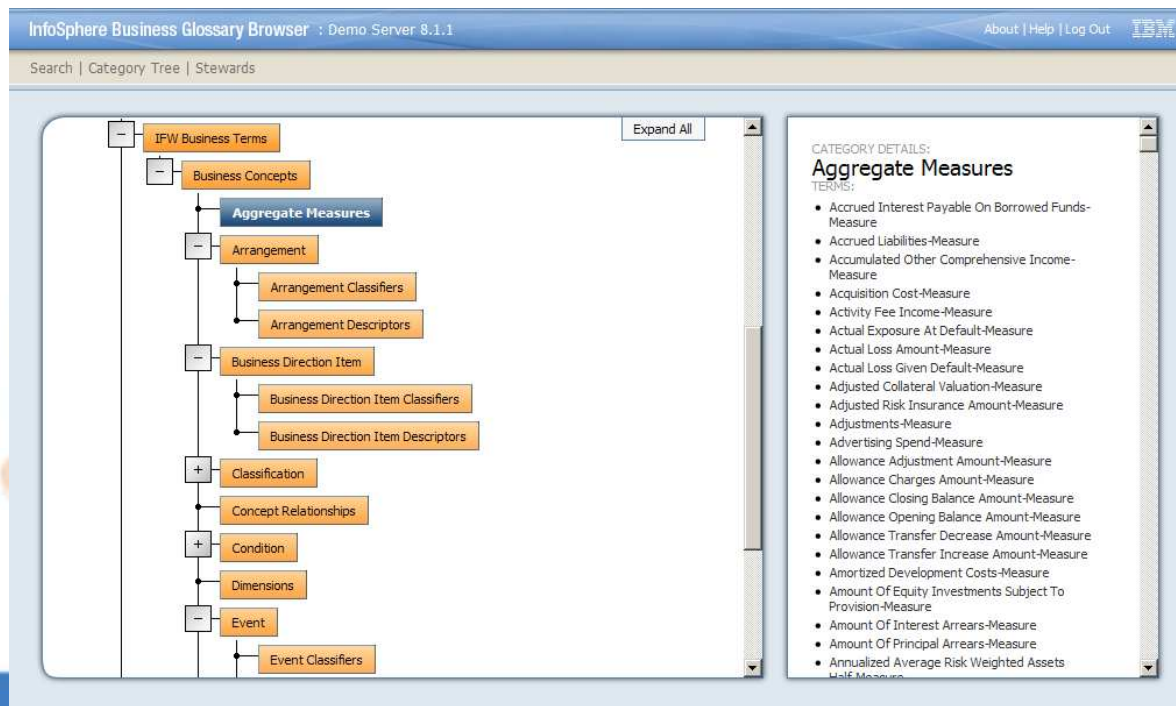


# InfoSphere Business Glossary



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

## Business Glossary



InfoSphere Business Glossary Browser : Demo Server 8.1.1

Search | Category Tree | Stewards

Expand All

- IFW Business Terms
  - Business Concepts
    - Aggregate Measures
    - Arrangement
      - Arrangement Classifiers
      - Arrangement Descriptors
    - Business Direction Item
      - Business Direction Item Classifiers
      - Business Direction Item Descriptors
    - Classification
    - Concept Relationships
    - Condition
    - Dimensions
    - Event
      - Event Classifiers

CATEGORY DETAILS:  
**Aggregate Measures**  
 TERMS:

- Accrued Interest Payable On Borrowed Funds-Measure
- Accrued Liabilities-Measure
- Accumulated Other Comprehensive Income-Measure
- Acquisition Cost-Measure
- Activity Fee Income-Measure
- Actual Exposure At Default-Measure
- Actual Loss Amount-Measure
- Actual Loss Given Default-Measure
- Adjusted Collateral Valuation-Measure
- Adjusted Risk Insurance Amount-Measure
- Adjustments-Measure
- Advertising Spend-Measure
- Allowance Adjustment Amount-Measure
- Allowance Charges Amount-Measure
- Allowance Closing Balance Amount-Measure
- Allowance Opening Balance Amount-Measure
- Allowance Transfer Decrease Amount-Measure
- Allowance Transfer Increase Amount-Measure
- Amortized Development Costs-Measure
- Amount Of Equity Investments Subject To Provision-Measure
- Amount Of Interest Arrears-Measure
- Amount Of Principal Arrears-Measure
- Annualized Average Risk Weighted Assets

## 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

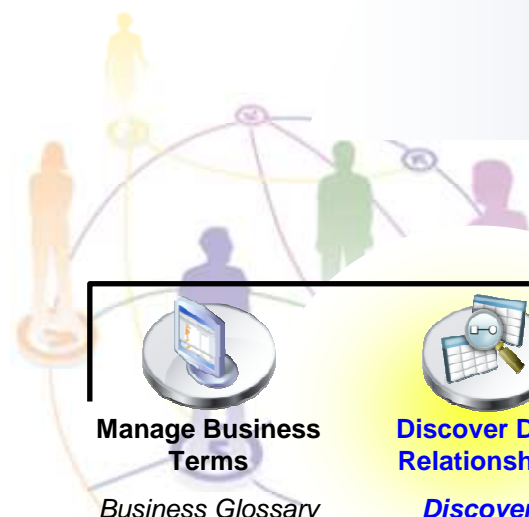


# InfoSphere Foundation Tools

## Enterprise Projects



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



<b>Manage Business Terms</b> <i>Business Glossary</i>	<b>Discover Data Relationships</b> <i>Discovery</i>	<b>Design Enterprise Models</b> <i>Data Architect</i>	<b>Capture Design Specifications</b> <i>FastTrack</i>	<b>Assess, Monitor, Manage Data Quality</b> <i>Information Analyzer</i>	<b>Monitor Data Flows</b> <i>Metadata Workbench</i>
--	--	--	--	--	--



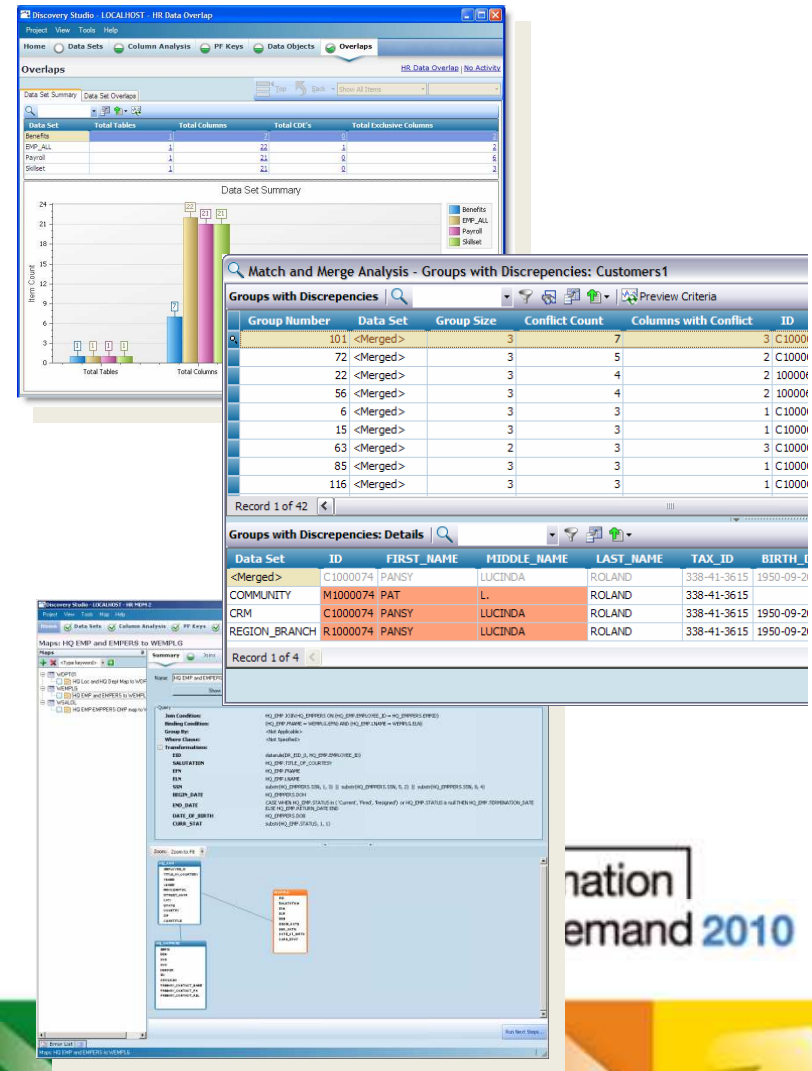
# InfoSphere Discovery

✓ Leverage unique capabilities of Discovery for rapid development

- **Automate data relationship analysis** and share with Information Analyzer for quality programs and Data Architect for enhanced enterprise modeling
- **Migrate off legacy applications** easier - discover transformation logic and generate replacement ETL jobs with FastTrack
- Leverage prototype consolidation capabilities with DataStage/QualityStage – **get it right the 1<sup>st</sup> time**
- **Create business objects** (logical units such as “customer”) for test data management, data growth and application retirement projects with Optim portfolio

✓ **Extended heterogeneous connectivity**

- DB2 zOS; IMS, VSAM, ADABAS, DataCom, IDMS, zOS flat files through Classic Federation Server



The screenshots show the Discovery Studio interface with the following components:

- Data Set Summary:** A table showing data sets and their attributes.
 

Data Set	Total Tables	Total Columns	Total CDE's	Total Exclusive Columns
EMP_ALL	1	22	1	1
Payroll	1	24	0	1
Subset	1	22	0	1
- Match and Merge Analysis - Groups with Discrepancies: Customers1:** A table listing groups with discrepancies.
 

Group Number	Data Set	Group Size	Conflict Count	Columns with Conflict	ID
101	<Merged>	3	7	3	C10000
72	<Merged>	3	5	2	C10000
22	<Merged>	3	4	2	100000
56	<Merged>	3	4	2	100000
6	<Merged>	3	3	1	C10000
15	<Merged>	3	3	1	C10000
63	<Merged>	2	3	3	C10000
85	<Merged>	3	3	1	C10000
116	<Merged>	3	3	1	C10000
- Groups with Discrepancies: Details:** A table showing detailed data for a specific group.
 

Data Set	ID	FIRST_NAME	MIDDLE_NAME	LAST_NAME	TAX_ID	BIRTH_DATE
<Merged>	C1000074	PANSY	LUCINDA	ROLAND	338-41-3615	1950-09-2
COMMUNITY	M1000074	PAT	L.	ROLAND	338-41-3615	
CRM	C1000074	PANSY	LUCINDA	ROLAND	338-41-3615	1950-09-2
REGION_BRANCH	R1000074	PANSY	LUCINDA	ROLAND	338-41-3615	1950-09-2
- SQL Query:** A window showing the generated SQL query for the data details.
 

```

SELECT EMP_FIRST_NAME, EMP_MIDDLE_NAME, EMP_LAST_NAME, EMP_TAX_ID, EMP_BIRTH_DATE
FROM EMP_ALL
WHERE EMP_ID = 'C1000074'

```

# InfoSphere Discovery



Discovery

Accelerate project deployment by automating discovery of your distributed data landscape

The screenshot shows the Discovery Studio interface with a data model diagram and a table of results.

Expression	Row Hit Rate	Value Hit Rate	Selectivity	Notes
ALL_POSITIONS_F->ALL_ORGANIZATION_UNITS				
ALL_ORGANIZATION_UNITS.ORGANIZATION_ID = ALL_POSITIONS_F.BUSINESS_GROUP_ID	2% (42/2235)	100% (3645/3645)	2% (42/2235)	100% (42/42)
ALL_ORGANIZATION_UNITS.ORGANIZATION_ID = ALL_POSITIONS_F.ORGANIZATION_ID	48% (1069/2235)	100% (3645/3645)	48% (1069/2235)	100% (1069/1069)

## Requirements

- Discover data transformation rules and heterogeneous relationships
- Identify hidden sensitive data for privacy
- Define business objects for archival, test data and other integration applications

## Benefits

- Automation of manual activities accelerates time to value
- Business insight into data relationships reduces project risk
- Provides consistency across information agenda projects

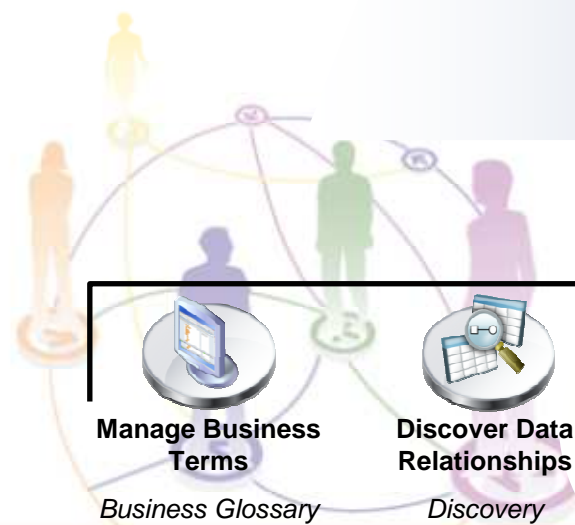


# InfoSphere Foundation Tools

## Enterprise Projects



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



<b>Manage Business Terms</b> <i>Business Glossary</i>	<b>Discover Data Relationships</b> <i>Discovery</i>	<b>Design Enterprise Models</b> <i>Data Architect</i>	<b>Capture Design Specifications</b> <i>FastTrack</i>	<b>Assess, Monitor, Manage Data Quality</b> <i>Information Analyzer</i>	<b>Monitor Data Flows</b> <i>Metadata Workbench</i>
--	--	--	--	--	--



# IBM Industry Models



IBM Industry Models

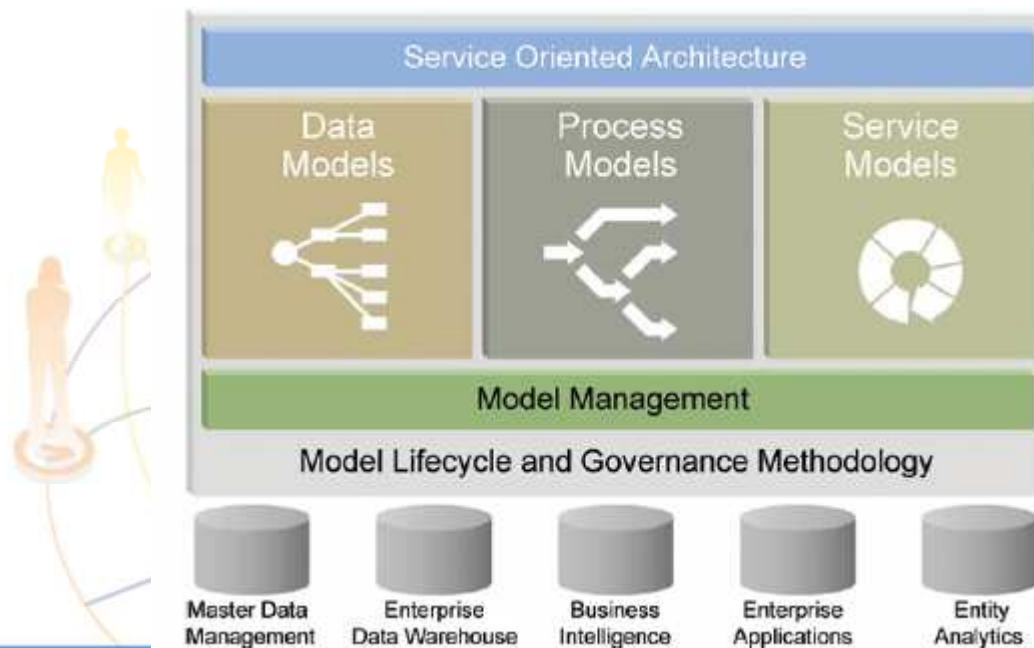
Leverage industry standard practices and processes, models and terminology to accelerate projects

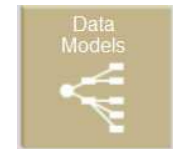
## Requirements

- Leverage industry standard practices and processes
- Populate Business Glossary with full industry vocabulary
- Accelerate data modeling activities for target systems

## Benefits

- Accelerate your Information Agenda
- Optimize business around key indicators
- Streamline workflows with industry templates





# IBM Industry Data Models Address the Following Business Areas

## **Banking** (Banking Data Warehouse)

- Profitability
- Relationship Marketing
- Risk Management
- Asset and Liability Mgmt
- Compliance

## **Financial Markets** (Financial Markets Data Warehouse)

- Risk Management
- Asset and Liability Mgmt
- Compliance

## **Health Plan** (Health Plan Data Warehouse)

- Claims
- Medical Management
- Provider and Network
- Sales, Marketing and Membership
- Financials

## **Insurance** (Insurance Information Warehouse)

- Customer centricity
- Claims
- Operational efficiency
- Underwriting; Pricing; Profitability
- Intermediary Performance
- Compliance
- Risk Management

## **Retail** (Retail Data Warehouse)

- Customer centricity
- Merchandising Management
- Store Operations & Product Mgmt
- Supply Chain Management
- Compliance

## **Telco** (Telecommunications Data Warehouse)

- Churn Management
- Relationship Mgmt & Segmentation
- Sales and Marketing
- Service Quality & Product Lifecycle
- Usage Profile

# InfoSphere 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



Data Analysts



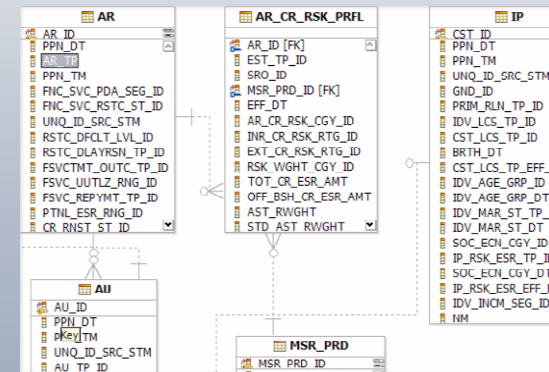
Architects

Design & Govern



Data Architect

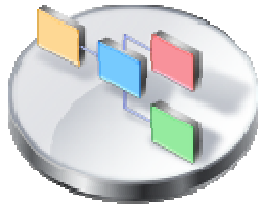
Model, visualize, and relate diverse and distributed data assets



Data Modeling, Mapping & Reconciliation

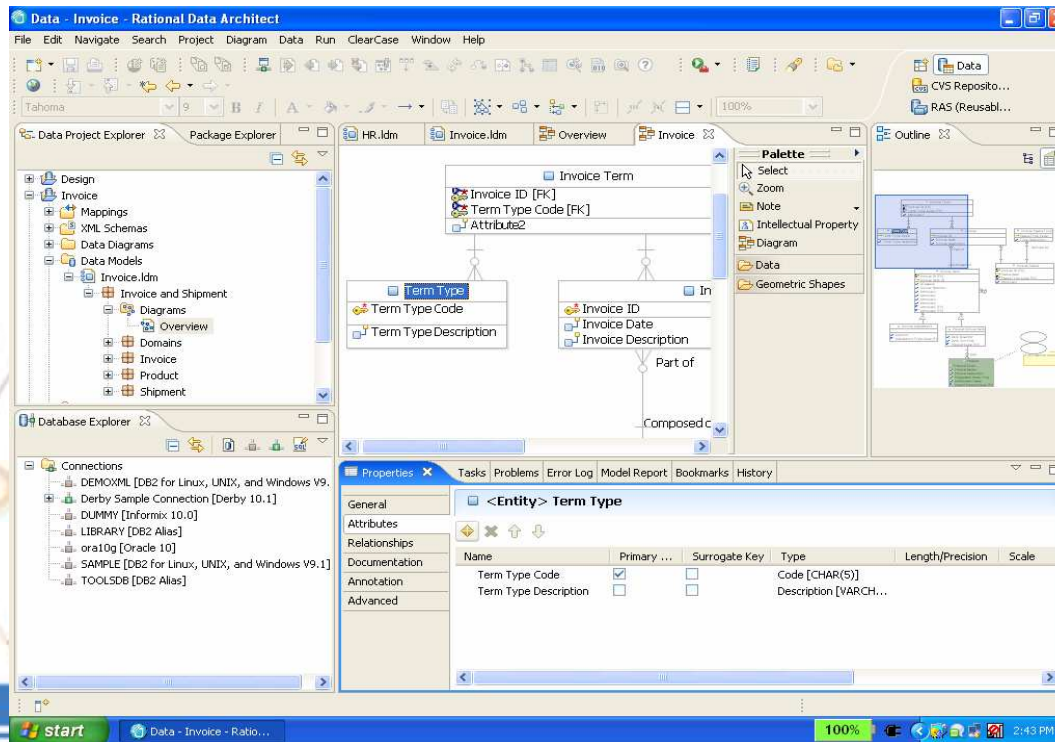


# InfoSphere Data Architect



**Data Architect**

**Model, visualize, and relate diverse and distributed data assets**



## Requirements

- Design and manage enterprise models
- Enforce model conformance to enterprise standards
- Leverage industry data models for best practices

## Benefits

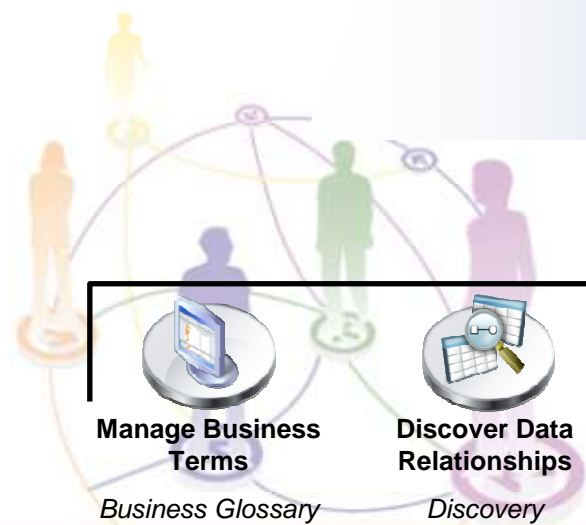
- Speed design activities
- Populate Business Glossary from model terms
- Validate models for enterprise conformance

# InfoSphere Foundation Tools

## Enterprise Projects



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

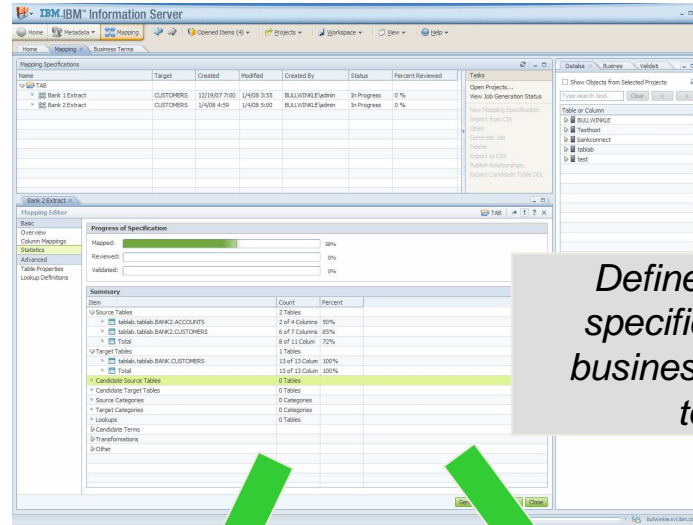


<b>Manage Business Terms</b>	<b>Discover Data Relationships</b>	<b>Design Enterprise Models</b>	<b>Capture Design Specifications</b>	<b>Assess, Monitor, Manage Data Quality</b>	<b>Monitor Data Flows</b>
<i>Business Glossary</i>	<i>Discovery</i>	<i>Data Architect</i>	<i>FastTrack</i>	<i>Information Analyzer</i>	<i>Metadata Workbench</i>



# InfoSphere FastTrack

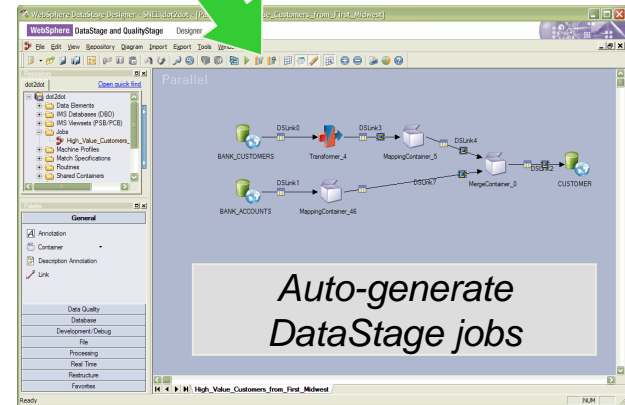
- Single managed infrastructure for data integration projects to track requirements to deployment
- Create required transformation rules for source-to-target mappings
- Define and link business terms to physical structures
- Generate DataStage jobs
- Generate historical documentation for tracking



Define mapping specification with business rules and terms



Flexible reporting and tracking



Auto-generate DataStage jobs

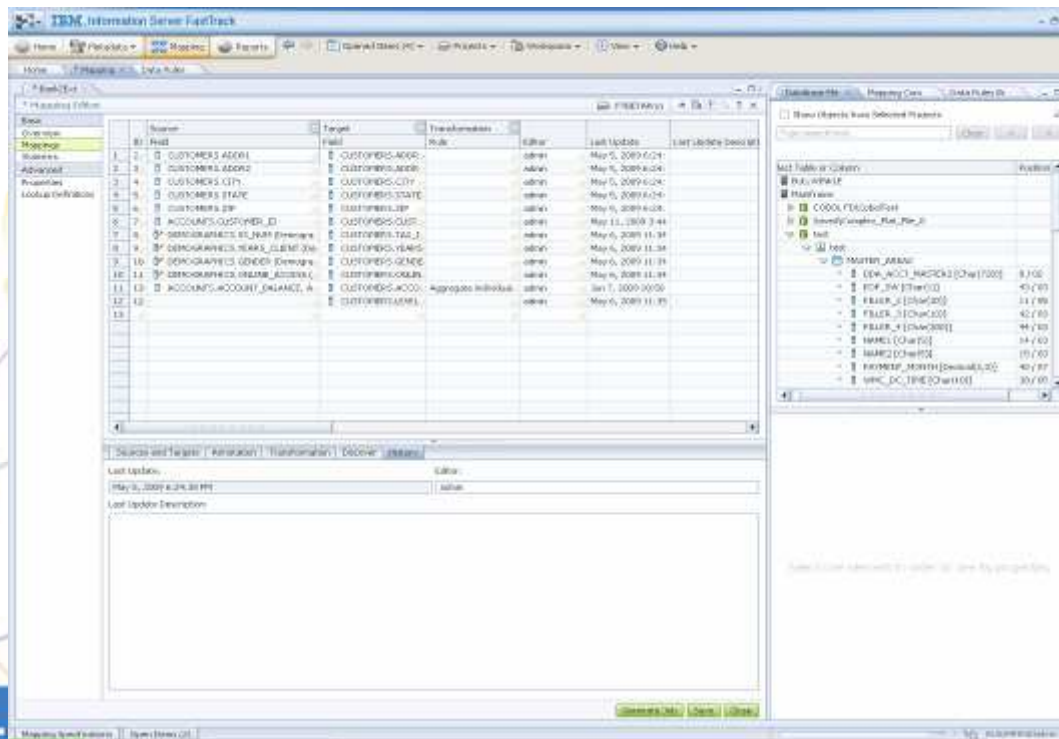


# InfoSphere FastTrack



FastTrack

Capture Design Specifications and accelerate translation into data integration projects



## Requirements

- Capture business requirements for source to target mappings
- Leverage source analysis and business vocabulary
- Generate candidate ETL jobs

## Benefits

- Accelerate development of integration processes
- Centralized management of specifications
- Audit design decisions over time

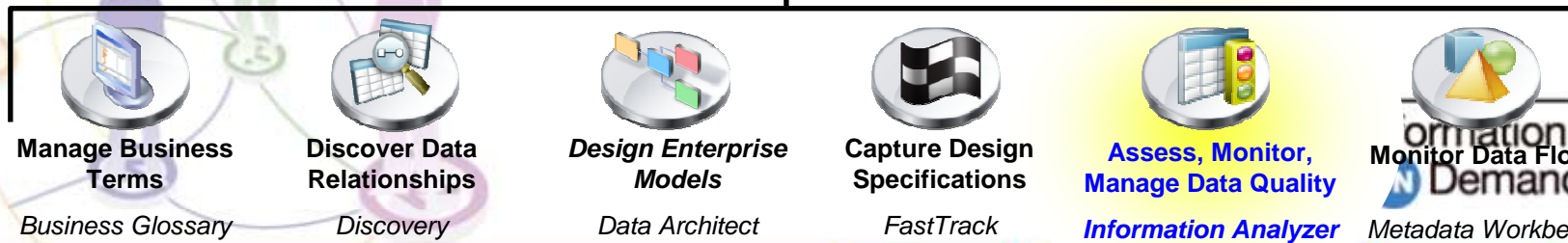
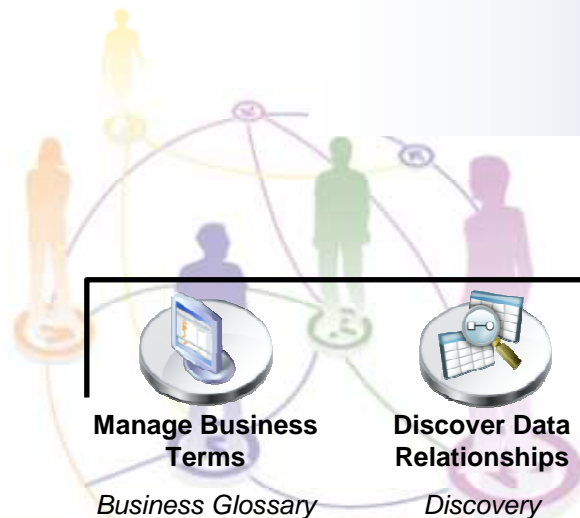


# InfoSphere Foundation Tools

## Enterprise Projects



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



Information Demand 2010



# InfoSphere Information Analyzer

- 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**

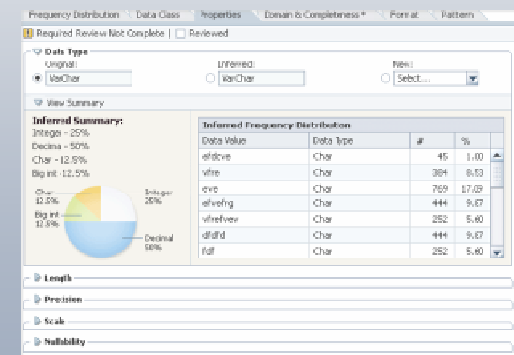


**Business & Data Analysts**

**Discover & Govern**



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



**Physical Understanding**

# Monitor Results over Time

- Data Quality Monitoring – scheduled processing

**Account Gender Exists**

Run

Job Name: \*  
Account Gender Exists run

Job Description:  
An execution of Account Gender

Run as Test

**Rule Executable**

Scheduler Sample Options

Run Now

Overview Result

Select View

By Record

By Distribution

By Rule

By Pattern

**Distribution by Rules Not Met**

View By: # of Rules Not Met

# of Rules Not Met	Run		Baseline	
	Record #	Record %	Record #	Record %
0	13894	94.7490 %	10966	74.7818 %
1	761	5.1896 %	3632	24.7681 %
2	9	0.0614 %	66	0.4501 %
3	0	0.0000 %	0	0.0000 %
4	0	0.0000 %	0	0.0000 %

**Baseline Set Comparison**

	Run	Baseline
Date/Time Executed	3/11/2009 12:45:50 PM	3/10/2009 4:57:4
Total Records	14664	14664
Mean Rules Not Met	1.3281 %	6.4171 %
Standard Deviation	5.6750 %	11.1747 %
Similarity	62.6696 %	
Degradation	0.0000 %	

**Carrier ID Validation**

View Results

Include Tests

**Job Log**

Type	Timestamp	Validity			Trend	Contact	Sample
		Severity	# Pass	# Fail			
Run	01/01/09	10.8%	792	208	⚠	Mim Foster	
Run	12/01/08	✓	904	96	⚠	Mim Foster	
Run	11/01/08	✓	926	74	⚠	Mim Foster	
Run	10/01/08	✓	500	0		Mim Foster	✓
Run	09/01/08	✓	500	0		Mim Foster	✓

**Records**

03% 06% 51%

% Rules Not Met

Run Baseline Similarity Degradation

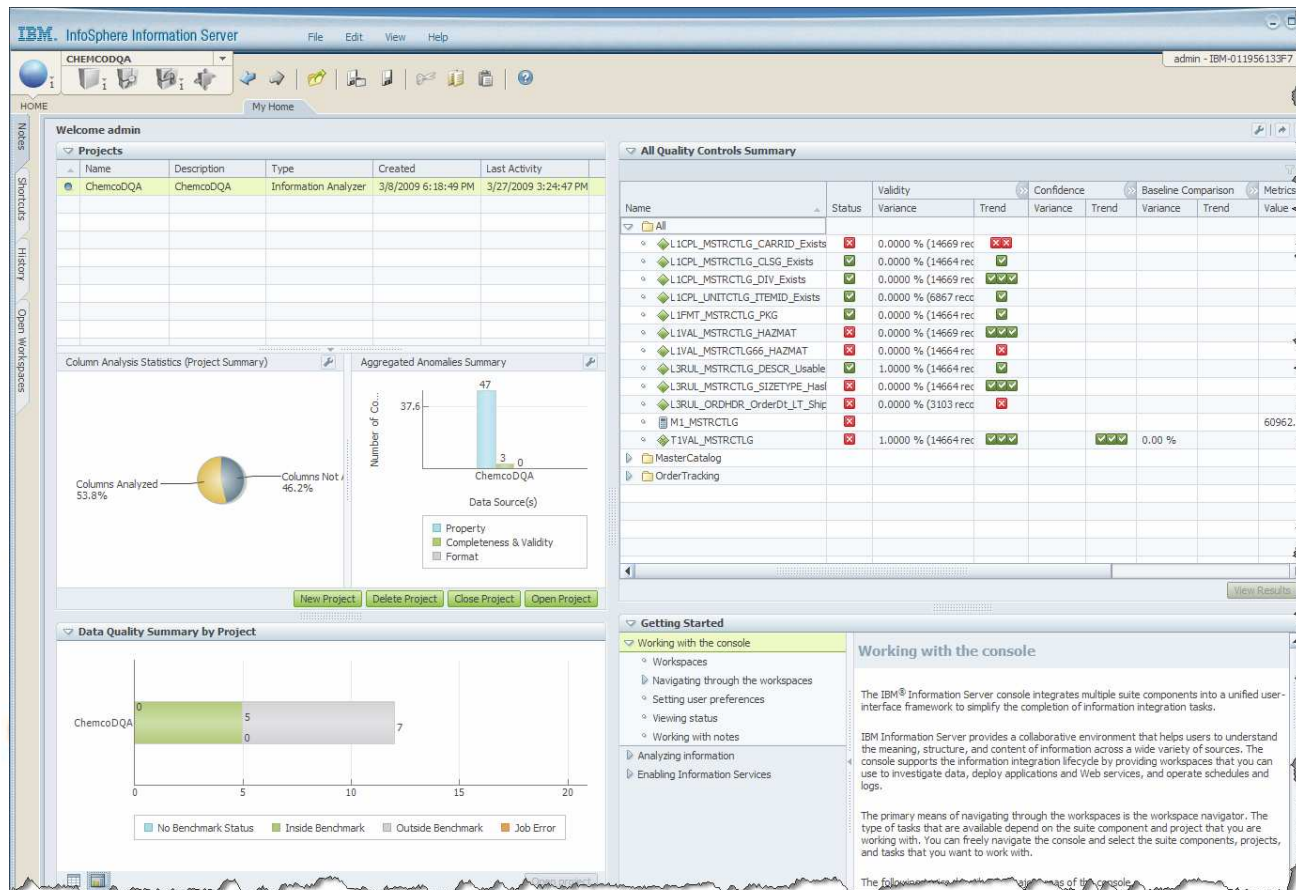
- Schedule & ad-hoc Data Rules, Rule Sets, and Metrics
- Annotate Execution History Events
- Analyze Specific Results, Trends, & Statistics
- View Execution Details



# View Summarized Results

- Dashboard Views & Alerts; Quality Reports

- Customize Data Quality Displays and Scorecards by project
- Organize by results by user-defined folders
- Display Standard Charts & Trends for Ongoing Views
- Track Analysis Status, Data Trends, Variances over Benchmarks, ...
- View Online Documentation & Screen Information





# InfoSphere Information Analyzer



Information Analyzer

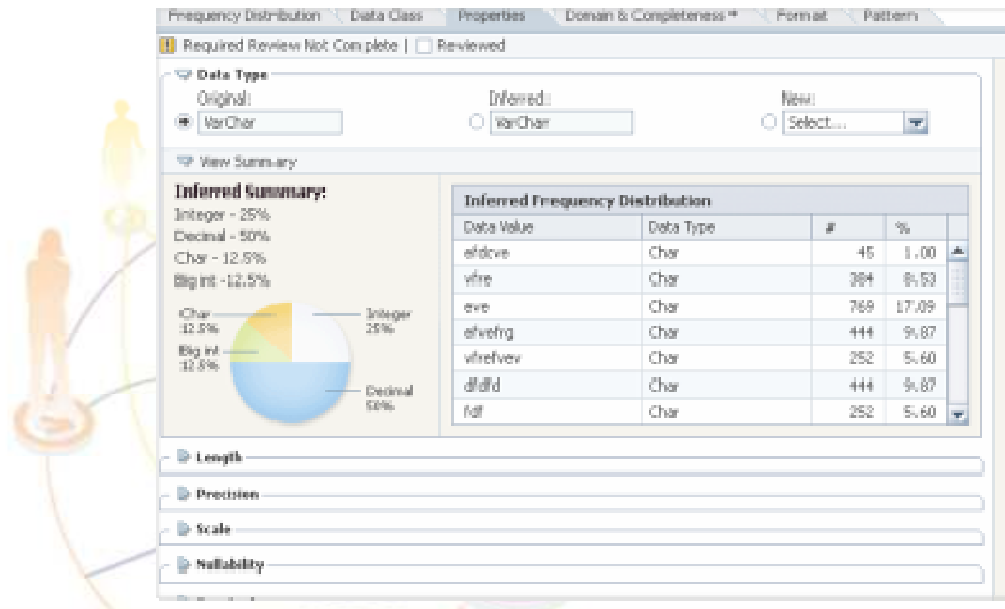
Analyze source data quality and monitor adherence to integration and quality rules

Requirements

- Perform data quality assessment
- Define business rules to monitor data quality
- Establish stewards for governance of data quality

Benefits

- Identify data quality issues early to reduce project risks
- Monitor quality metrics over time for compliance
- Create business confidence with trusted information

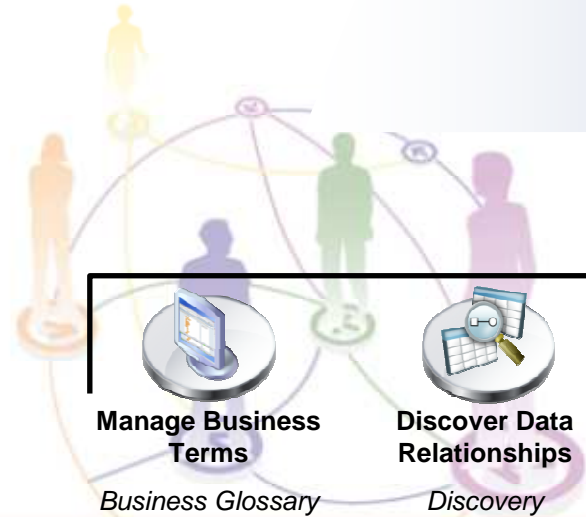


# InfoSphere Foundation Tools

## Enterprise Projects



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




<b>Manage Business Terms</b> <i>Business Glossary</i>	<b>Discover Data Relationships</b> <i>Discovery</i>	<b>Design Enterprise Models</b> <i>Data Architect</i>	<b>Capture Design Specifications</b> <i>FastTrack</i>	<b>Assess, Monitor, Manage Data Quality</b> <i>Information Analyzer</i>	<b>Monitor Data Flows</b> <i>Metadata</i>
--	--	--	--	--	--




# 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
  - Trace data lineage of modeling, BI reports and InfoSphere objects




**Subject Matter Experts**



**Governance Stewards**



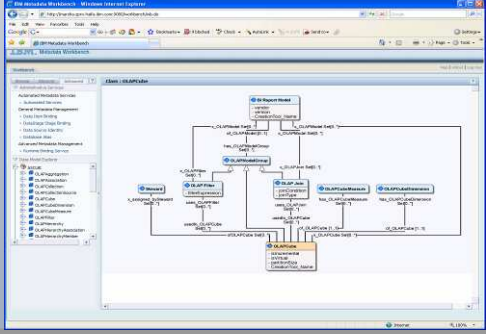
**Architects**



**Design & Govern**

**InfoSphere Metadata Workbench**

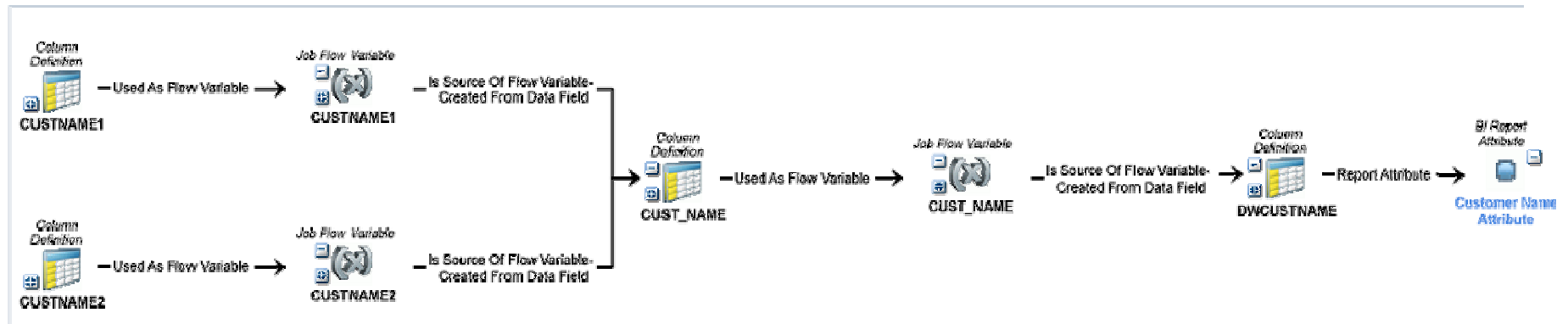
Support information governance with traceability on InfoSphere data movement, modeling & BI applications



**Impact Analysis & Lineage**

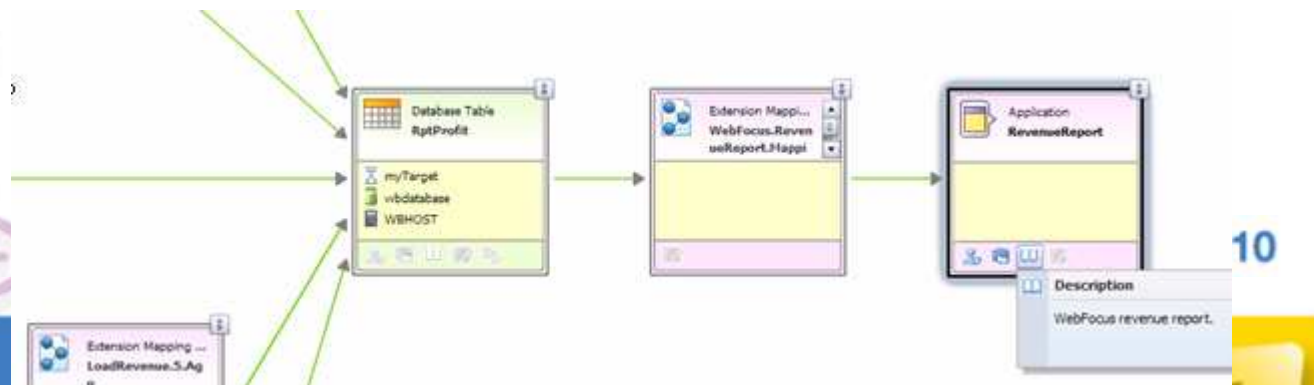
# Where does a Field of Data in this Report Come From?

Data Movement for: Customer Name Attribute



- We can trace data lineage beyond just BI tools.

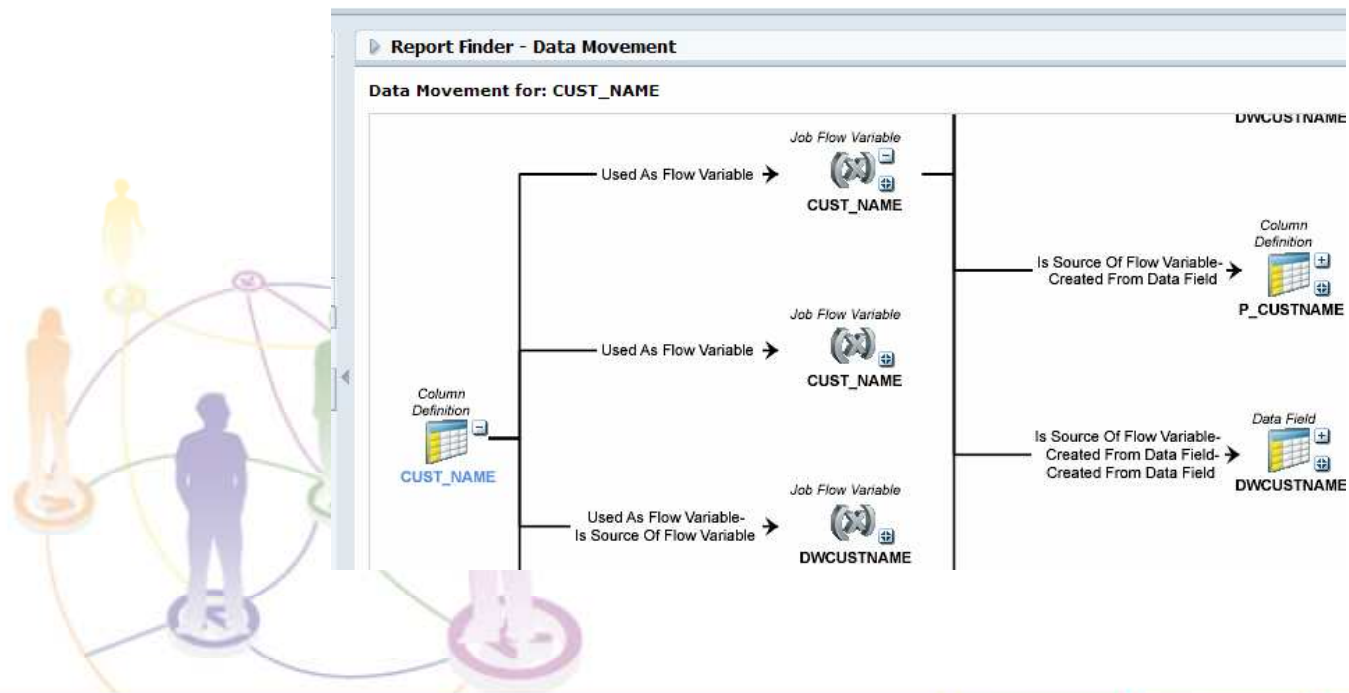
- FTP
- Stored Procedures
- MQ
- Anything....



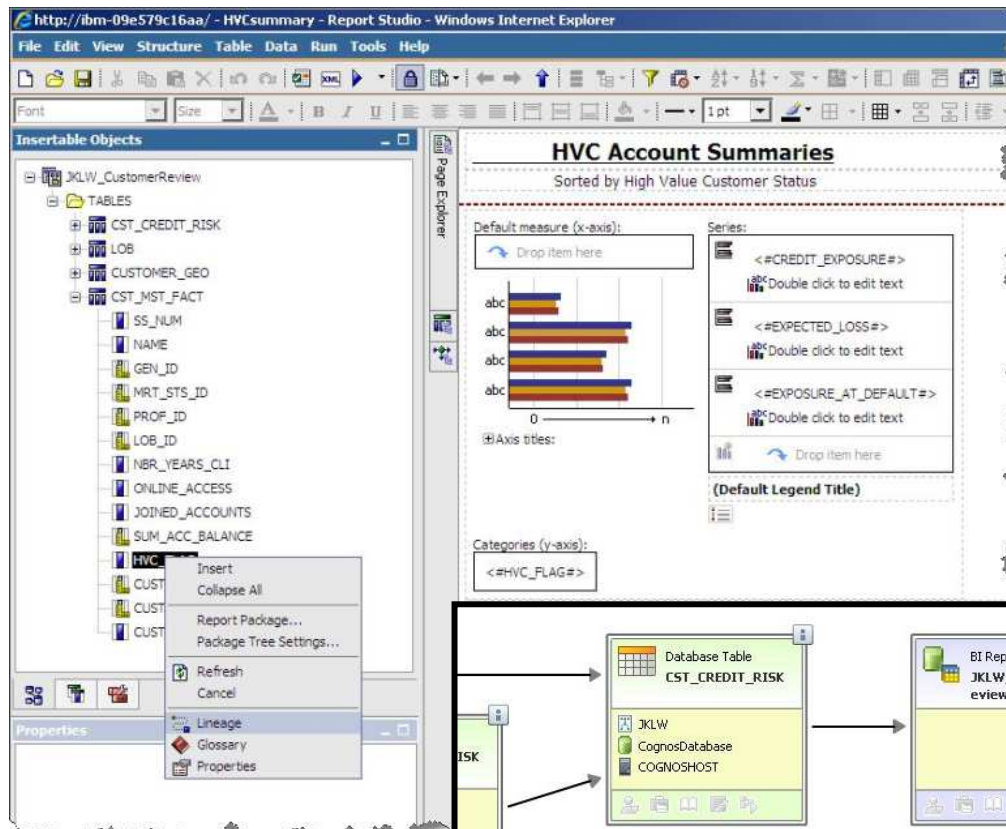


## Where Happens if I Change this Column?

- *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*

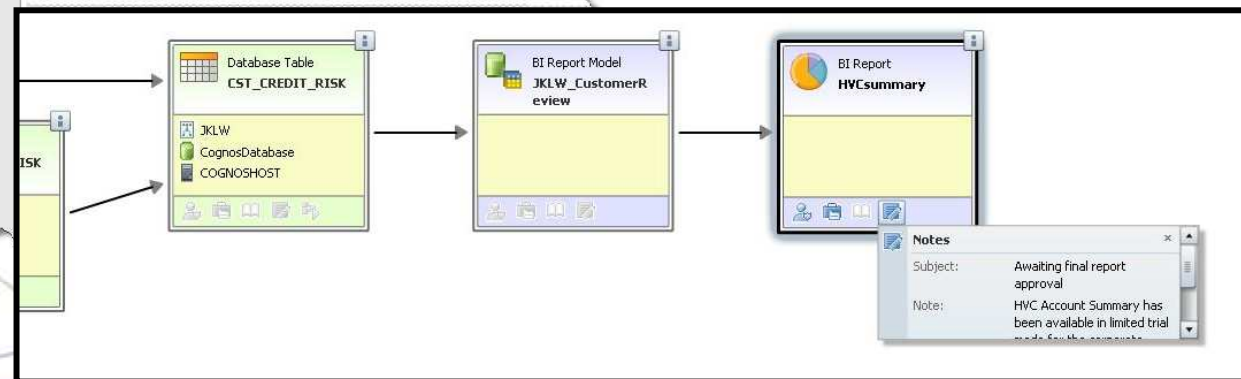


# Using Cognos to trace Data Lineage



We want to know

- where the data comes from
  - where it flows to
  - how it is transformed
- as it travels through the enterprise



# InfoSphere Metadata Workbench



Metadata Workbench

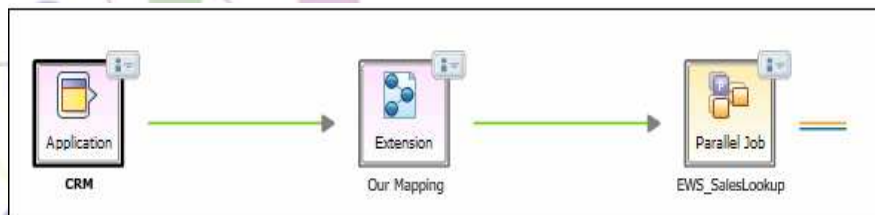
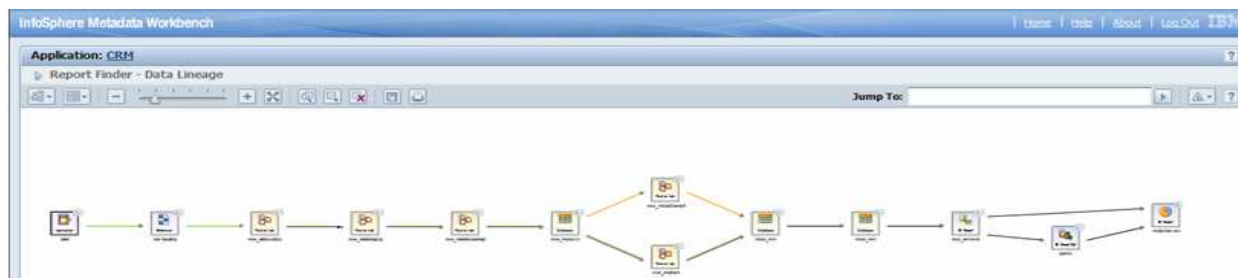
Support information governance with traceability on data movement, modeling & BI applications

## Requirements

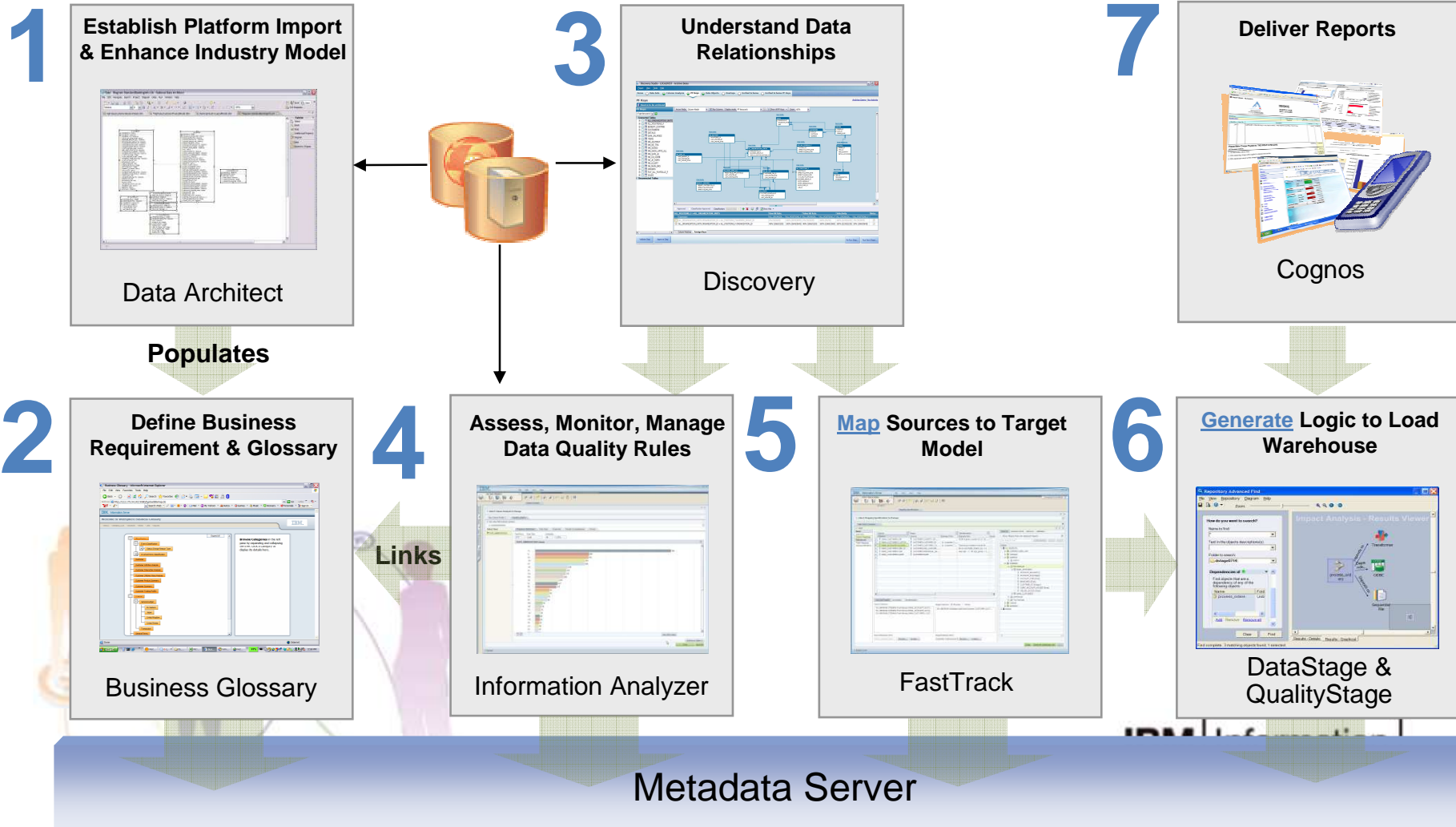
- Understand the impact of making changes to information environment
- Visualize and trace information flows across enterprise landscape
- Access and report on operational metadata

## Benefits

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



# Integrates Foundation Tools & Beyond



*Simplification & Content: reduces project time, risk and cost!*





# Your Choice...

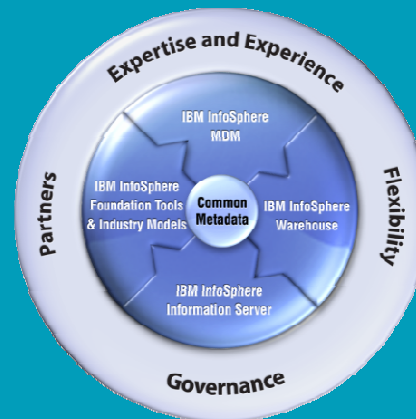
## Point Products

? + TRILLIUM SOFTWARE + INFORMatica + siperian + ORACLE + Business Objects + ?

Models	Cleansing	ETL	MDM	Warehouse	BI	Mashups
--------	-----------	-----	-----	-----------	----	---------

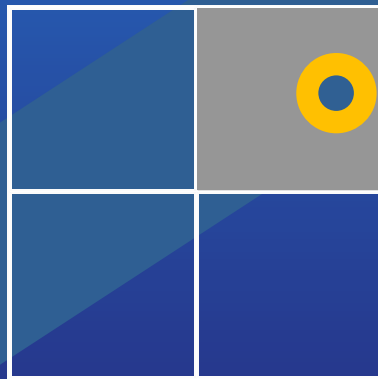
IBM + IBM + IBM + IBM + IBM + IBM + IBM

## Integrated Platform



# Leading the way

*Industry Analysts Rank IBM as Leaders....*



- *ETL Tools*
- *Business Intelligence Platforms*
- *Data Warehousing*
- *Data Quality Tools*
- *Data Integration Tools*
- *Customer Data Integration*

*and many other firms...*

**InfoSphere™**

Thank You!



IBM Information ON Demand 2010