



Governing Your Enterprise Data with IBM InfoSphere Foundation Tools

Jo Friedman
InfoSphere Product Manager
IBM Israel Software Lab

InformationOnDemandIndia2011

The Premier Conference for Information Management
Manage. Analyze. Govern.

February 2, 2011

Hyatt Regency | Mumbai, India

Disclaimer



© Copyright IBM Corporation 2010. All rights reserved.

U.S. Government Users Restricted Rights - Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

Information regarding potential future products is intended to outline our general product direction and it should not be relied on in making a purchasing decision. The information mentioned regarding potential future products is not a commitment, promise, or legal obligation to deliver any material, code or functionality. Information about potential future products may not be incorporated into any contract. The development, release, and timing of any future features or functionality described for our products remains at our sole discretion.

IBM, the IBM logo, ibm.com, InfoSphere, InfoSphere Business Glossary, InfoSphere Data Architect, InfoSphere Information Analyzer, InfoSphere FastTrack, and InfoSphere Metadata Workbench are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. If these and other IBM trademarked terms are marked on their first occurrence in this information with a trademark symbol (® or ™), these symbols indicate U.S. registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at “Copyright and trademark information” at <http://www.ibm.com/legal/copytrade.shtml>

Other company, product, or service names may be trademarks or service marks of others.



Agenda



- What is the problem?
- What is Information Governance and how is it important to my business?
- IBM and Information Governance
- Getting Started
- InfoSphere Foundation Tools



Challenges around governing the usage, sharing and processing of massive amounts of electronic data



Lack of trusted information

- No alignment of definitions across business and IT
- No clear understanding of data sources & relationships
- No standardized quality rules or threshold metrics
- Lack of control over test data environments

Growing infrastructure and resource management costs

- No policies for management of data growth – performance degradation
- Redundancy of data
- Disparate, complex applications and more users

Lack of common security and privacy approach

- Risk of security breaches, compliance, audit failures
- No ability to assess areas of vulnerability and prevent unauthorized intrusion
- Lack of an overall protection strategy (relational / non-relational data and access controls)

“By 2013, 25% of the companies in highly regulated industries will create and staff positions in accounting, human resources, compliance and audit and law that deal explicitly with the management of information via technology.”

– **Gartner, Inc.**, “*Organizing for Information Governance*”
Debra Logan, November 2009

“...An [Information Management] strategy should incorporate life-cycle information governance practices...[for] business optimization, agility, and transformation.

– **Forrester Research**, “*Refresh Your Information Management Strategy to Deliver Business Results*”
Rob Karel & James G. Kobielus, August 2009

“If you are going to protect your company's most valuable asset—your data—you will begin to view data security as a component of a more comprehensive information governance strategy.”

– **Hurwitz & Associates**, “*Why you need an information governance strategy for 2010*”
Marcia Kaufman, December 2009

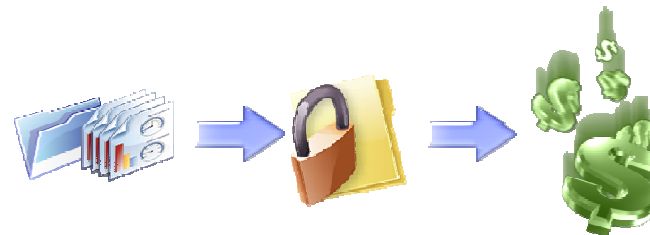
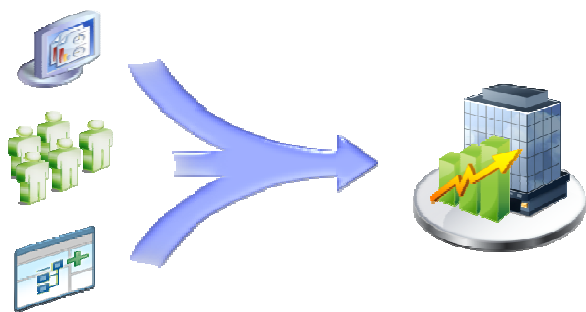




Information Governance Creates Order out of Information Chaos

Information Governance is the exercise of decision rights to optimize, secure and leverage data as an enterprise asset.

- Orchestrate people, process and technology toward a common goal
- Promotes collaboration
- Derive maximum value from information
- Leverage data as an enterprise asset to drive opportunities
- Safeguards information
- Ensure highest quality
- Manage it throughout lifecycle



Governing the creation, management and usage of enterprise data is not an option any longer. It is:

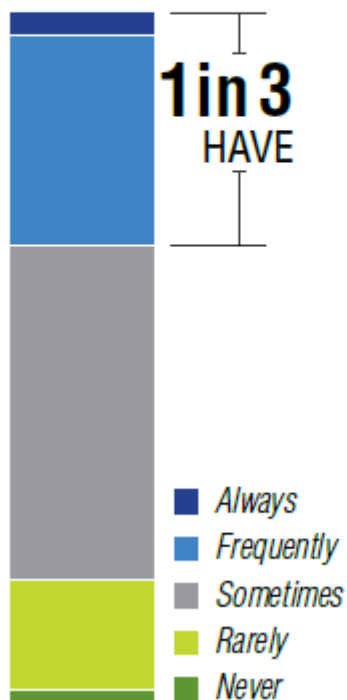
Expected by your customers ♦ Demanded by the executives ♦ Enforced by regulators/auditors



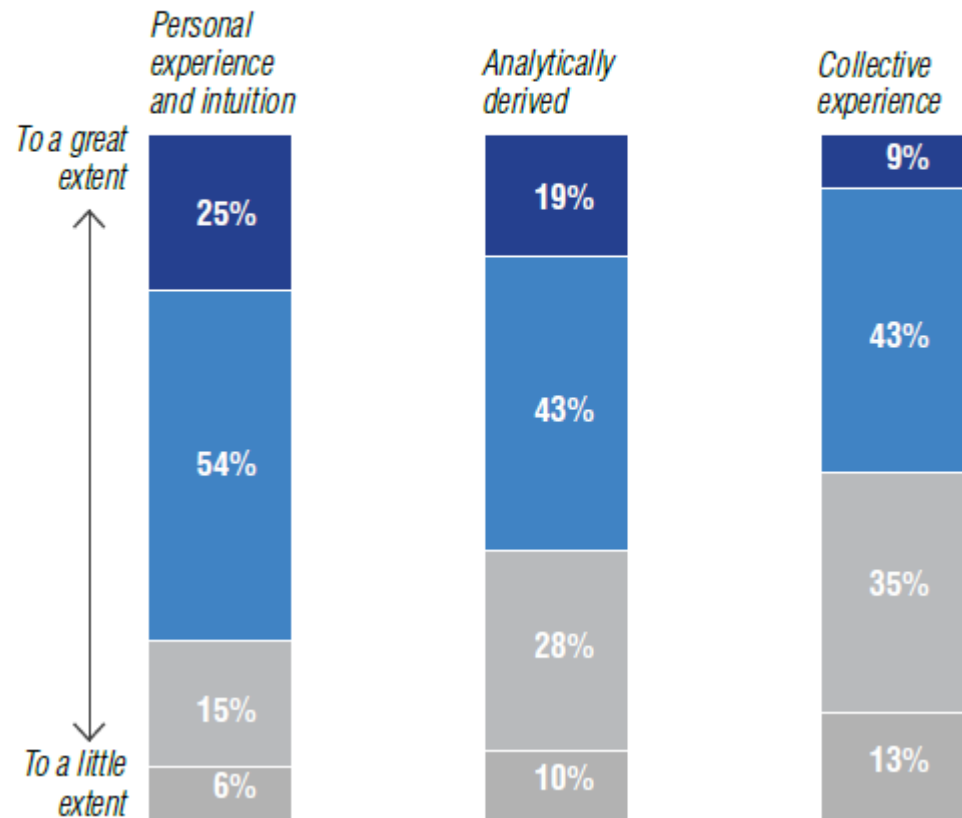


Information Governance: Why? ...Pain Points

How often have you made major decisions with incomplete information or information you don't trust?

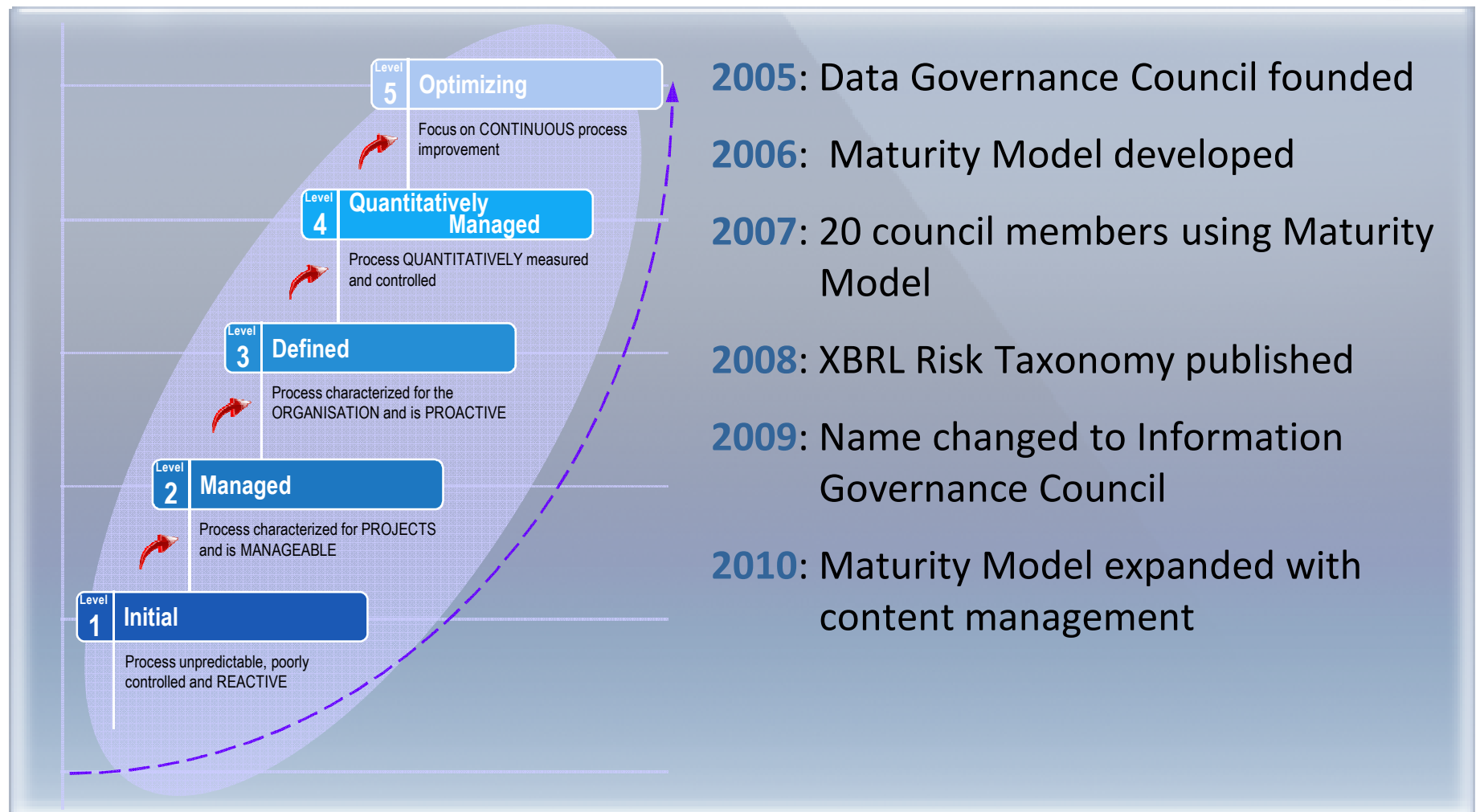


To what extent do you make business decisions based on the following factors?





IBM's History of Thought Leadership in Information Governance



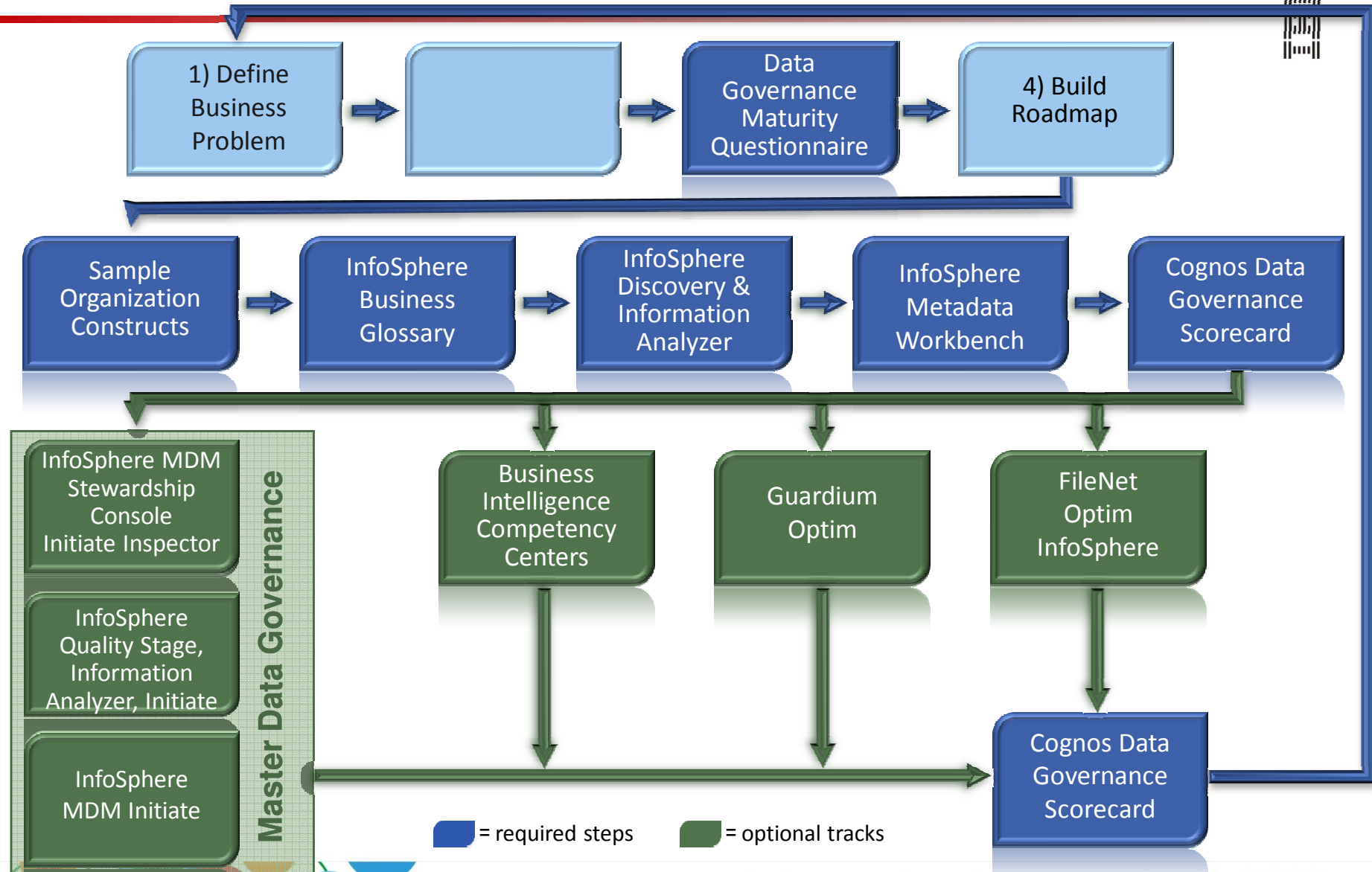


Information Governance Council Maturity Model



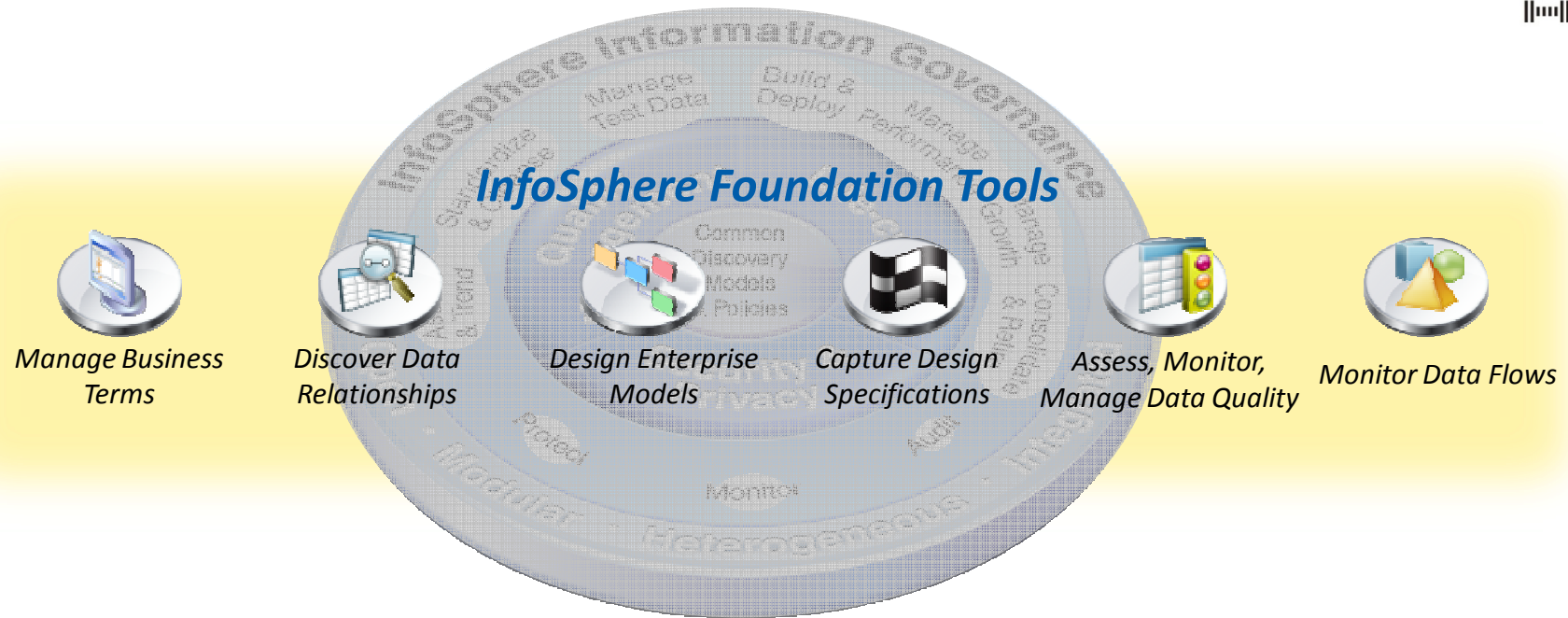


IBM Data Governance Unified Process



InfoSphere Foundation Tools

Put your arms around your data ... for Trusted Information



- Jump start any data integration, MDM, data warehousing, business intelligence or information governance project
- Discover your data across systems
- Design your trusted information structure
- Govern your information over time

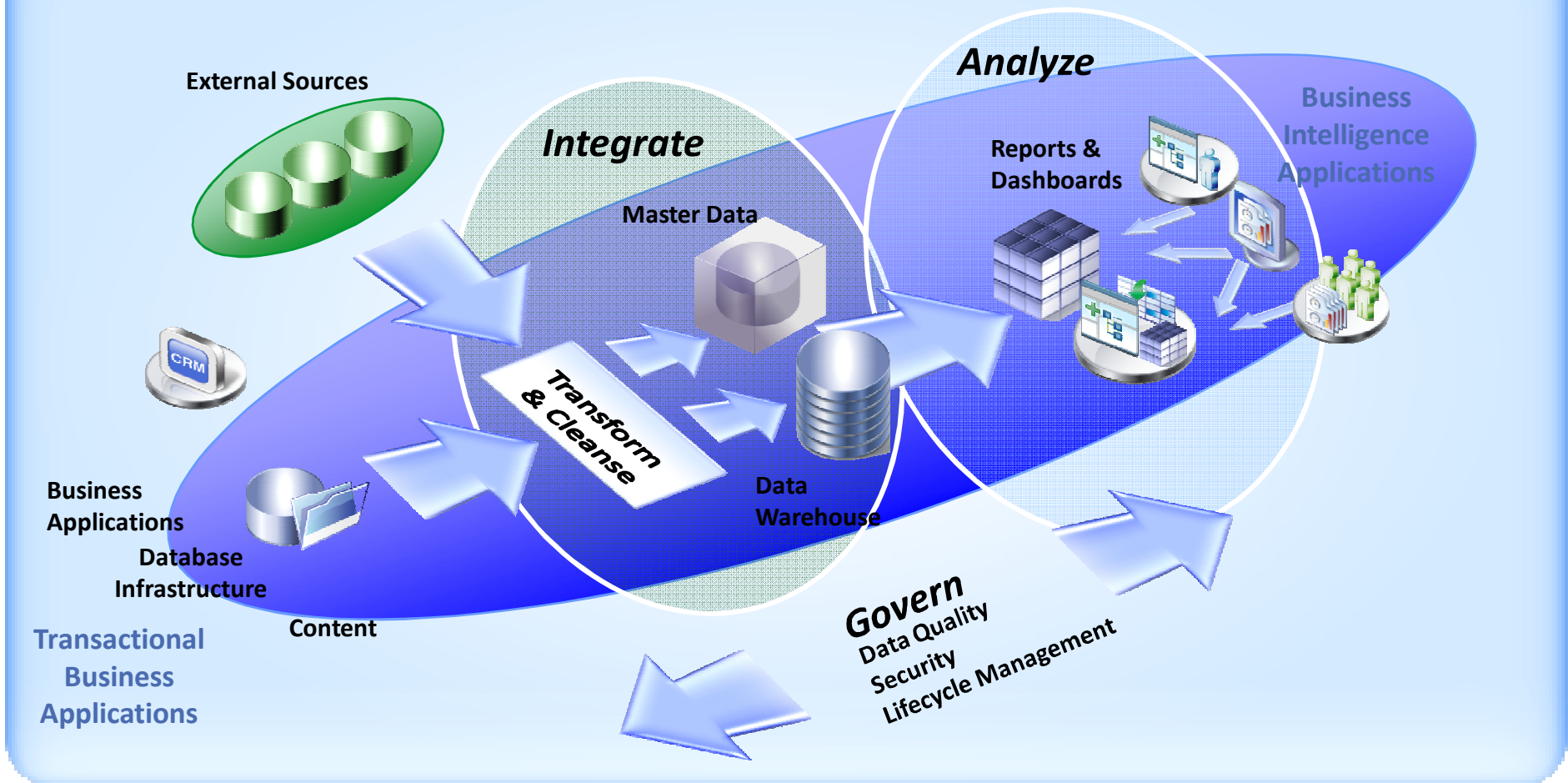


Delivering business insight requires mastering your information

Going beyond simply managing... to Optimizing Business Results



Your Information Supply Chain





Information Governance and InfoSphere Foundation Tools

Key Governance Capability

- Establish a common set of terms and definitions
- Understand the data I have in my organization
- Understand the way that data flows and how it is connected
- Assign meaning to data assets
- Assign ownership to data assets
- Monitor data quality

Foundation Tools product

- ✓ **InfoSphere Business Glossary**
- ✓ **InfoSphere Discovery & InfoSphere Information Analyzer**
- ✓ **InfoSphere Metadata Workbench**
- ✓ **InfoSphere Business Glossary, Information Analyzer, Metadata Workbench**
- ✓ **InfoSphere Business Glossary, InfoSphere Metadata Workbench**

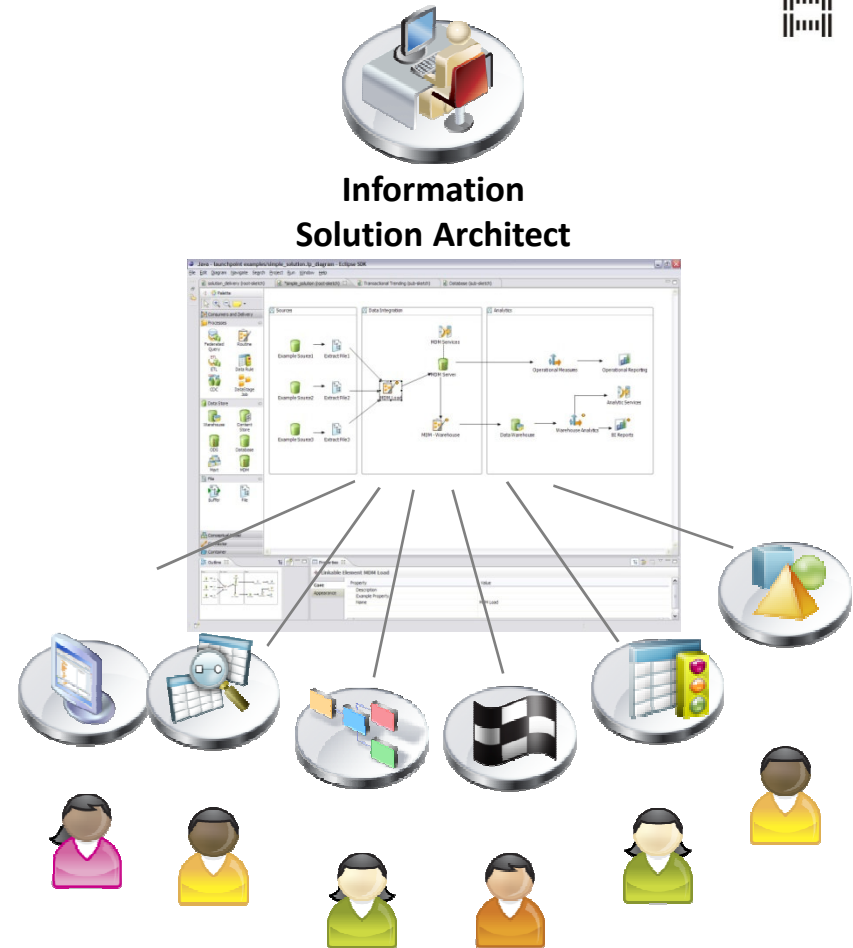


Blueprint Director

The GPS for your Information Project



- Create and manage actionable blueprints of your information project
- Accelerate information centric projects by leveraging templates with methods – e.g. warehousing, master data – and customize your best practices
- Navigate through your information project by managing the information roadmap & its evolution over time
- Establish business-driven development by
 - aligning business & IT views
 - creating a consistent end-to-end design from business requirements



Information project leadership team
(stakeholders, bus. Analysts, stewards, specialty architects, etc.)

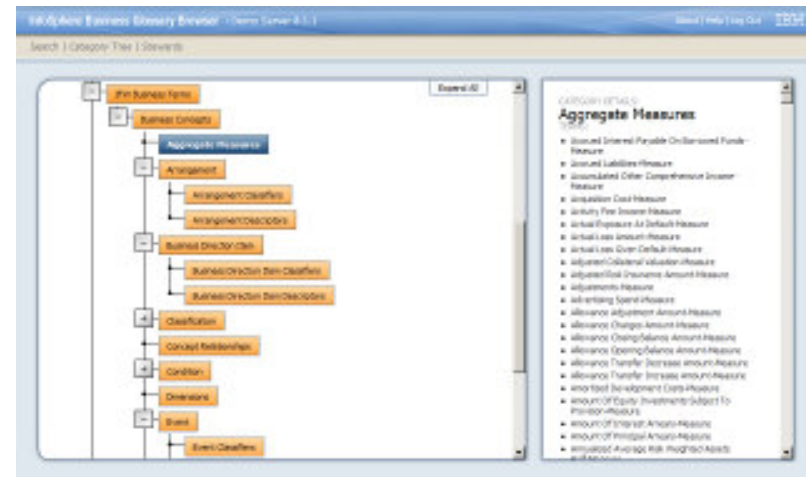




Business Glossary: Define your information and its meaning

Capabilities:

- Capture important terms e.g. for key data elements
- Classify your assets: where do I find which business term
- Assign data stewards: who is responsible for which data domain (and terms)
- Benefits
 - Align business & IT
 - Faster time to value
 - Reduce risk
 - Establish your governance strategy



InfoSphere Business Glossary Browser - Welcome to InfoSphere Business Glossary

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-FIIIIVVVV.
Status: Standard

General Information

Assigned Assets

- Customers_Table (Local Host > Northwind_Database > Northwind_Schema)
- Marketing Campaign Email Message TYPE: WEM Task CONTEXT: Marketing Campaign >> Mailing Members Process LOCATION: NE Process Server Node
The task mails a campaign letter to golden club members.
- SimpleReport1

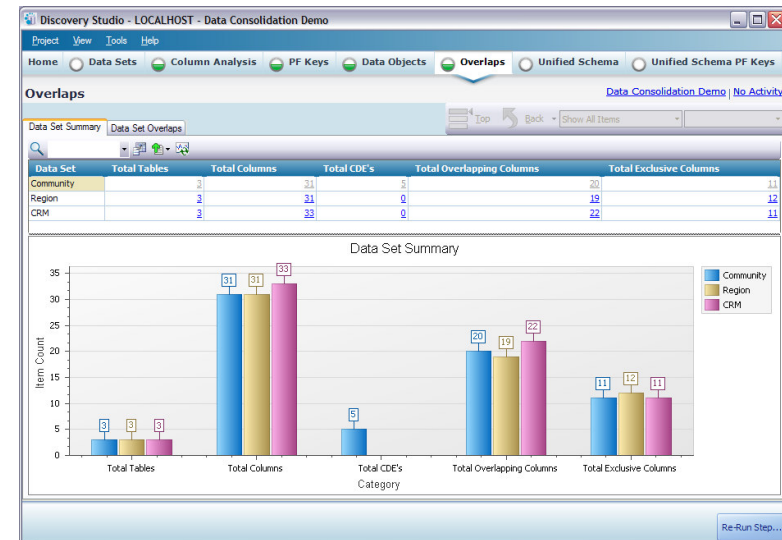
Notes
History



InfoSphere Discovery: Discover your information



- Capabilities
- Automatic data analysis
- Identify data relationships; e.g. customer, counterparty, invoice
- Relate source data objects across multiple sources
- Identify technical data elements to populate the hub
- Benefits
 - Increase your time to value by automating the analysis of your data sources



The screenshot shows the 'Column Summary' section of the Discovery Studio interface for 'HR Data Overlap'. It displays a detailed table of column overlaps across three data sets: EMP_ALL, Benefits, and Payroll. The table includes columns for Data Set, Table, Column, Column Number, CDE, Cardinality, Selectivity, Non Nulls %, and Value Overlap with Benefits, Payroll, and Sidset.

Summary	Data Set	Table	Column	Column Number	CDE	Cardinality	Selectivity	Non Nulls %	Value Overlap with Benefits	Value Overlap with Payroll	Value Overlap with Sidset
EMP_ALL	EMP_ALL	EMP_ID	1	50	1.00	100%	0.0%	100%	82%	100%	24%
EMP_ALL	EMP_ALL	DEPT_ID	2	28	0.96	100%	0.0%	100%	75%	90%	25%
EMP_ALL	EMP_ALL	TITLE	3	20	0.40	100%	0.0%	100%	75%	75%	50%
EMP_ALL	EMP_ALL	MANAGER_ID	4	6	0.12	100%	0.0%	100%	83%	12%	83%
EMP_ALL	EMP_ALL	FIRST_NAME	5	29	0.98	100%	0.0%	100%	84%	100%	24%
EMP_ALL	EMP_ALL	LAST_NAME	6	30	1.00	100%	0.0%	100%	20%	85%	24%
EMP_ALL	EMP_ALL	ADDRESS_1	7	50	1.00	100%	0.0%	100%	70%	85%	24%
EMP_ALL	EMP_ALL	ADDRESS_2	8	1	0.02	100%	0.0%	100%	0%	0%	0%
EMP_ALL	EMP_ALL	ADDRESS_CITY	9	50	1.00	100%	0.0%	100%	82%	100%	24%
EMP_ALL	EMP_ALL	ADDRESS_STATE	10	22	0.58	100%	0.0%	100%	90%	100%	34%
EMP_ALL	EMP_ALL	ADDRESS_COUNTRY	11	44	0.88	100%	0.0%	100%	84%	100%	27%
EMP_ALL	EMP_ALL	ADDRESS_ZIP	12	50	1.00	100%	0.0%	100%	82%	100%	24%
EMP_ALL	EMP_ALL	PHONE_HOME	13	20	1.00	100%	0.0%	100%	82%	100%	24%
EMP_ALL	EMP_ALL	PHONE_CELL	14	20	1.00	100%	0.0%	100%	0%	0%	24%
EMP_ALL	EMP_ALL	PHONE_WORK	15	30	1.00	100%	0.0%	100%	0%	0%	24%
EMP_ALL	EMP_ALL	EC_NAME	16	28	0.96	100%	0.0%	100%	0%	0%	24%
EMP_ALL	EMP_ALL	EC_PHONE_1	17	50	1.00	100%	0.0%	100%	0%	0%	24%
EMP_ALL	EMP_ALL	DATE_HIRED	18	50	1.00	100%	82%	100%	82%	100%	24%
EMP_ALL	EMP_ALL	EMPLOYEE_BY	19	23	0.46	100%	0.0%	100%	83%	85%	33%
EMP_ALL	EMP_ALL		20	1	0.01	100%	0.0%	100%	0%	0%	0%

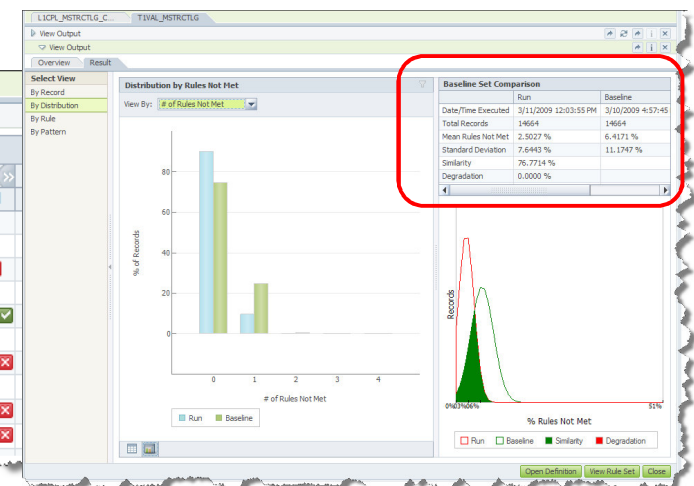




Information Analyzer: Data Quality Monitoring

- **Integrated Rules Analysis** to provide development, deployment, and evaluation of critical data rules on an ongoing basis
- **Business-driven Rule Definition** with reusable construction and application across multiple data sources speeding time-to-value
- **Multi-level Rule Assessment of Data Sources** with evaluation at the rule-, record-, and source-levels providing significantly greater insight into potential quality issues
- **Full, Sampled, or Stratified Source Processing** through user-driven configuration allowing for flexible data quality evaluations
- **Adhoc, Scheduled, or Command-driven Execution** with parallel processing support drives data quality monitoring when and where it is needed

Data Quality Alerts											
Name	Status	Total Records	Meet All Rules		Did Not Meet 1 c		% of Rules Not Met Per Record		Validity		
			#	%	#	%	Mean	Standard Deviation	Variance	Trend	
MasterCatalog											
L1CPL_MSTRCTLG_CARRID_Exists	✗	14669	14461	98.5820 %	208	1.4180 %			1.4180 %	✗✗	
L1CPL_MSTRCTLG_CLSG_Exists	✓	14664	14664	100.0000 %	0	0.0000 %			0.0000 %	✓	
L1CPL_MSTRCTLG_DIV_Exists	✓	14669	14669	100.0000 %	0	0.0000 %			0.0000 %	✓✓✓	
L1FMT_MSTRCTLG_PKG	✓	14664	14664	100.0000 %	0	0.0000 %			0.0000 %	✓	
L1VAL_MSTRCTLG_HAZMAT	✗	14669	14663	99.9591 %	6	0.0409 %			0.0409 %	✗✗✗	
L3RUL_MSTRCTLG_DESCR_Usable	✓	14664	14587	99.4749 %	77	0.5251 %				✓	
L3RUL_MSTRCTLG_SIZETYPE_HashChar	✗	14664	14096	96.1266 %	568	3.8734 %			3.8734 %	✗✗✗	
T1VAL_MSTRCTLG	✗	14664	13894	94.7490 %	770	5.2510 %	1.3281 %	5.6750 %	623.36664	✗✗✗	





Information Analyzer: Data Quality Reporting

- **Dashboard and Reporting Capabilities** in support of rules monitoring
- **80+ out-of-the-box Reports** viewable through IA UI or Reporting Console browser in XML, HTML, PDF, RTF, or TXT format
- **Command line Execution** providing additional integration approaches for delivery of data quality information

The screenshot displays the Information Analyzer interface with several key components:

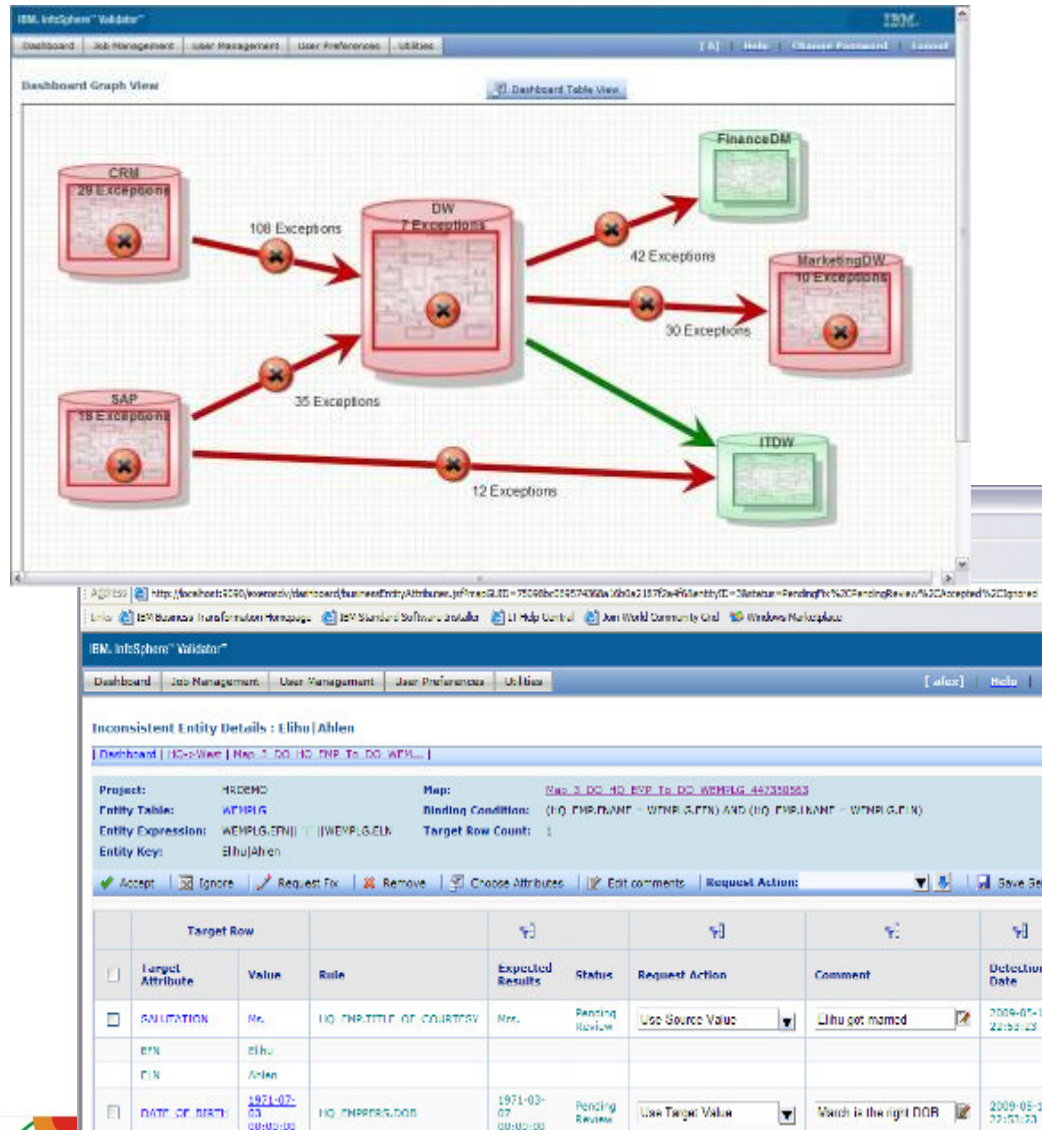
- Left Navigation Panel:** Lists various report categories such as Information Analyzer, Baseline Analysis, Column Classification, Column Domain, Column Frequency, Column Inferred, and Column Properties.
- Table:** A table with columns for Name and Description, listing various reports like 'Baseline Content - Current to Prior Variances' and 'Column Classification'.
- Column Level Summary:** A table showing summary statistics:

Format Cardinality Count	10
Value Cardinality Count	4738
Null Count	112567
Total Actual Row Count	824634
- Most Frequent Value Count:** A bar chart showing the count of values for different categories. The Y-axis is labeled 'Count' and ranges from 0 to 700,000. The X-axis shows categories like '9999', '999', '9999-', 'AAAAA', '9999', '9999A', '99', '9', '9999', and 'NA'. The 'NA' category has the highest count, exceeding 600,000.
- Projects Table:** A table with columns for Name, Description, Type, Created, and Last Activity. It lists a project named 'ChemcoDQA'.
- All Quality Controls Summary:** A table listing quality control rules with columns for Name, Validity, Variance, Trend, and Confidence. It includes rules like 'L1CPL_MSTRCTLG_CAR' and 'L1CPL_MSTRCTLG_CLSK'.
- Column Analysis Statistics (Project Summary):** A pie chart showing 'Columns Analyzed 53.9%' and 'Columns Not Analyzed 46.2%'. A bar chart shows the 'Number of Columns' for 'ChemcoDQA' with a value of 47.
- Aggregated Anomalies Summary:** A bar chart showing the number of anomalies for 'ChemcoDQA' with a value of 3.
- Bottom Panel:** Includes a 'Welcome admin' message, a 'Projects' table, and a 'Column Analysis Statistics' section.

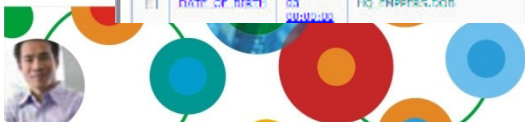




Exception Manager: Manage Data Quality Exceptions



- Ongoing validation and remediation of data quality rules
- Reconciliation of exceptions
- Previously found, fixed and ignored
- Browser based quality dashboard and remediation console for the Data Steward
- Drillable quality stats
- Configurable display
- Trending reports
- Excel report generation
- Exception lifecycle tracking





Metadata Workbench: Trace Data Movement

Capabilities

- Trace your data movement from sources to applications and business reports
- Understand relationships across your data flows
- Analyze the impact of changes to your data
- Centralized management of your IT assets

Benefits

- **Govern your data** by understanding where its coming from and where its being used
- **Support audit and regulatory compliance requirements** by reporting on your operational metadata

The screenshot displays the IBM Business Linkage Viewer interface. The top section shows a data flow diagram with nodes representing data sources and destinations. Below this, a detailed view of a 'DataStage Job' is shown, including its configuration and usage.

DataStage Job Configuration:

Property	Value
Job	SalesTransaction
Type	SERVER
Project	IAIRBIT-OTM3LITAN
Folder	\\nlb01\ACFM_Sales
Description	Load into all current transactional information into the operational data store, preserving activity of the past 12 months.
Data Steward	Jeddie Smith
Container	Container: DataStage - Blazoo
Contents	File Operational_Sales QueryDate Sales_Report Sales_Report_Source Reading_Sales_Figures Write Writing Sales Figures To DB Writing_Sales_Figures_To_DB
Container	Container: DataStage Local Container

DataStage Job Usage:

Operational Job Locator Information	SoftwareExecutable(Job) - SalesTransaction
Job Runs	SalesTransaction (2007-01-24 16:21:17) SalesTransaction (2007-01-24 18:26:44) SalesTransaction (2007-01-24 17:06:21)
Previous DataStage Job	Order: Entertainment





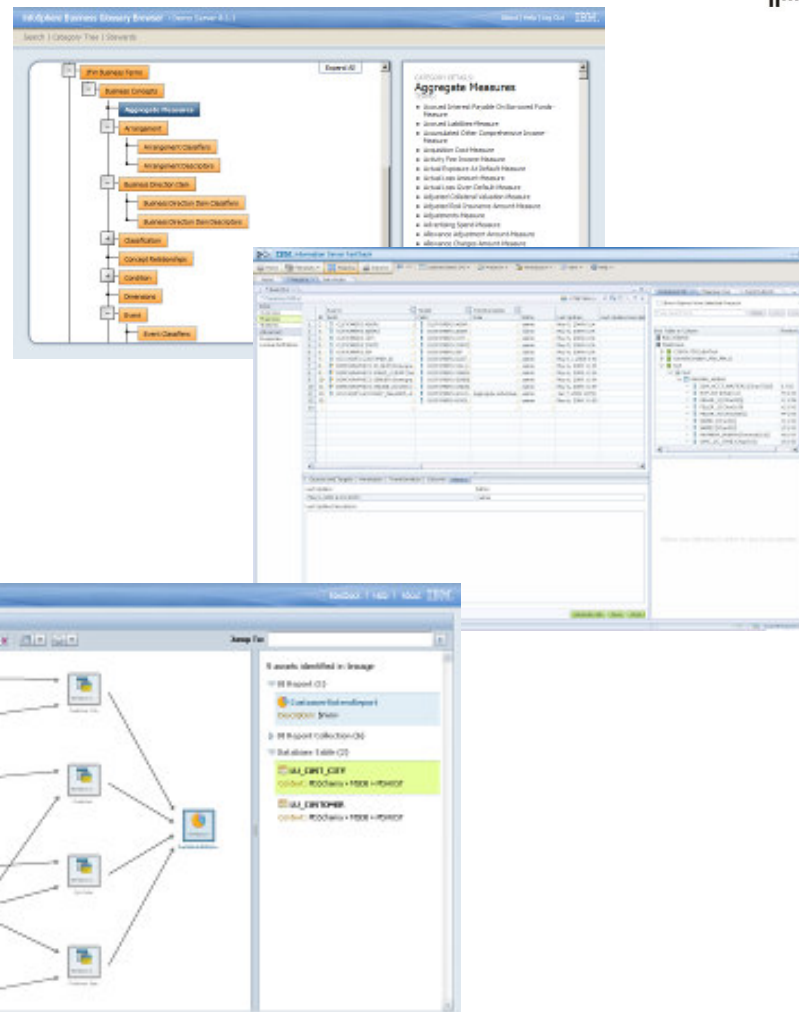
Business Glossary: Document

Capabilities

- Capture and communicate your terms and relationships to IT assets
- Document assignment of data stewards
- Record your mapping specifications
- Report on your end-to-end data flows

Benefits

- Governance
- Compliance
- Lasting Value





Information Governance Resources

Recommended reading:

IBM Data Governance Unified Process

Sunil Soares – Director, Information Governance

<https://i2.infoprint.com/sales/catalogs.nsf/agswgcatalogint?openagent>

IBM Data Governance Council

<http://www.ibm.com/ibm/servicemanagement/us/en/data-governance.html>





Thank You

InformationOnDemandIndia2011

The Premier Conference for Information Management
Manage. Analyze. Govern.

February 2, 2011

Hyatt Regency | Mumbai, India