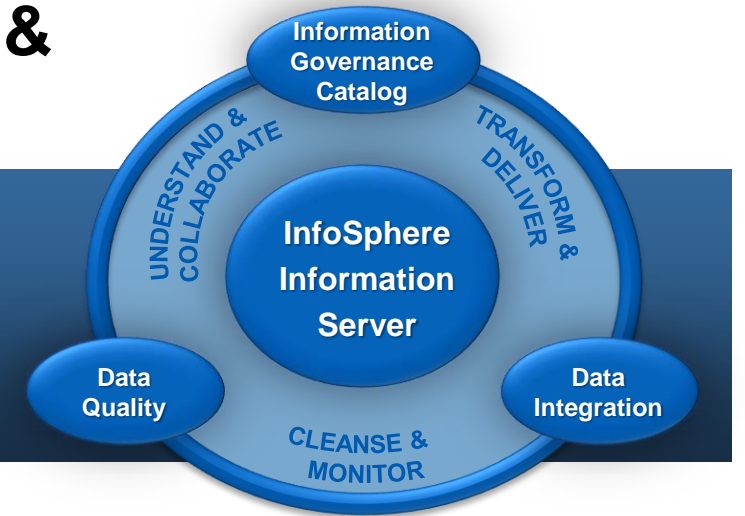


# Unified Platform for Data Integration, Data Quality & Data Governance

## ... Powered by IBM InfoSphere Information Server

Integrating and transforming data and content to deliver accurate, consistent, timely and complete information on a single platform unified by a common metadata layer



### Data Integration

#### *Transform & Deliver*

- Massive scalability
- Power for any complexity
- Deliver in batch and/or real-time with change capture



### Data Quality

#### *Cleanse & Monitor*

- Analyze, classify & validate
- Cleanse & standardize
- Define, manage & monitor data rules & exceptions



### Information Governance Catalog

#### *Understand & Collaborate*

- Catalog technical metadata & align w/ business language
- Manage (big) data lineage
- Manage information governance policies

- common connectivity • shared metadata • common execution engine
- flexible deployment: On-premise, Grid, Cluster, Cloud or native on Hadoop

# Unified Platform for Data Integration, Data Quality & Data Governance

## Information Server – Data Quality

Analyze, cleanse and monitor your data quality in a unified environment

**Most comprehensive data quality capabilities that run on premises or on cloud**

### ■ Analyze

- Discovers data of interest to the organization based on business defined data classes
- Analyzes data structure, content and quality
- Automates your data analysis process

### ■ Cleanse

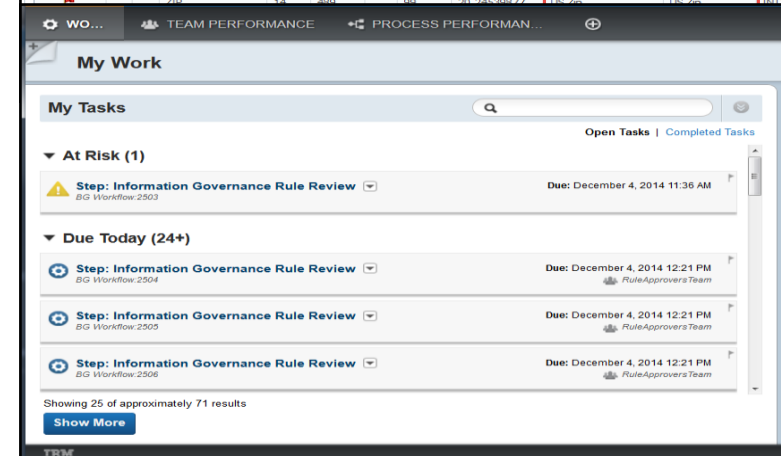
- Investigate, standardize, match and survive data at scale and with the full power of common data integration processes

### ■ Monitor

- Assess and monitor the quality of your data in any place and across systems
- Align quality indicators to business policies
- Engage data steward team when issues exceed thresholds of the business

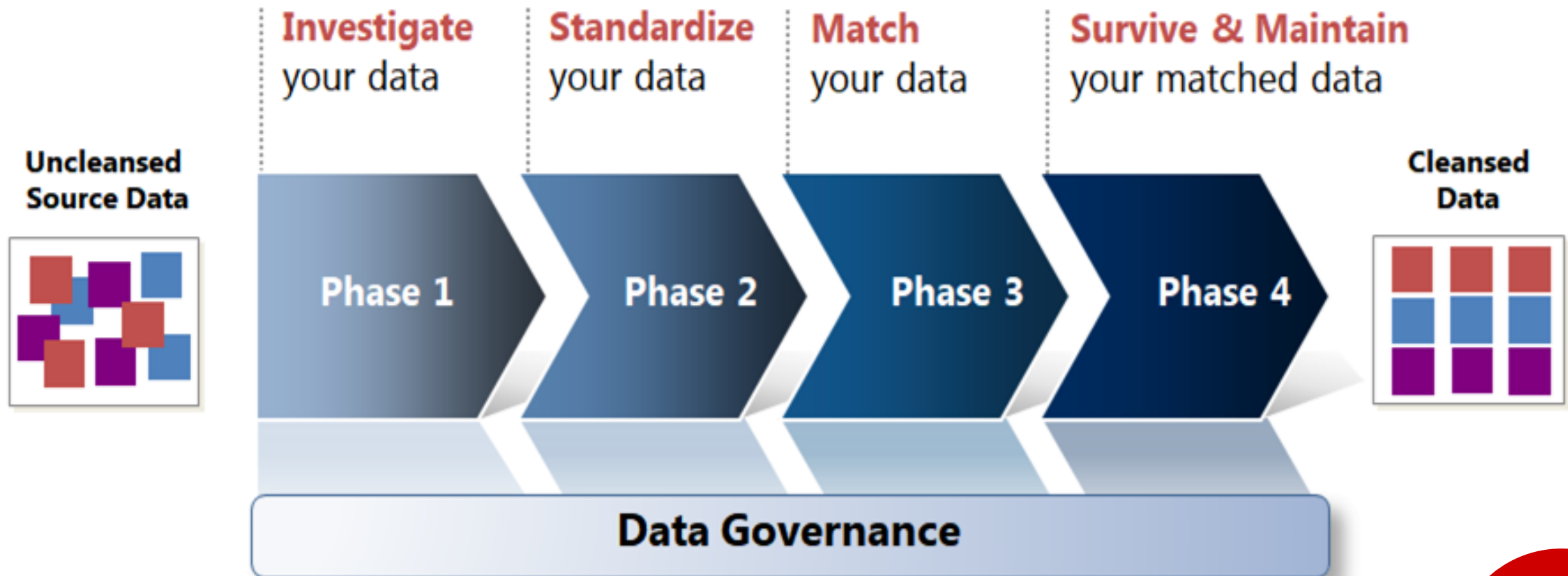


Table Totals									
Total Rows	489	Total Columns	25	Data Class	0	Properties	0	Domain	0
Column Attributes Reviewed									
BANK_ACCOUNTS: (25 of 25 columns)									
Name	Seq	Records	Defl	Cardinality	Inferred	Selected	Data Type	Length	Precision
CITY	1	489	70	14,314,928,43	Text	Text	STRING	16	Inferred
PROFESSION	2	489	18	3,680,981,60	Text	Text	STRING	24	Inferred
NAME	3	489	469	95,910,020,45	Text	Text	STRING	27	Inferred
ADDR1	4	489	459	93,865,030,67	Text	Text	STRING	31	Inferred
ADDR2	5	489	9	1,840,490,80	Text	Text	STRING	21	Inferred
STATE	6	489	9	1,840,490,80	Text	Text	STRING	2	Inferred
GENDER	7	489	2	0,408,997,96	Text	Text	STRING	1	Inferred
SAVINGS_ACCOUNT	8	489	3	0,613,496,93	Code	Indicator	STRING	3	Inferred
ONLINE_ACCESS	9	489	3	0,613,496,93	Boolean	Code	STRING	3	Inferred
JOINED_ACCOUNTS	10	489	3	0,613,496,93	Boolean	Code	STRING	3	Inferred
BANKCARD	11	489	3	0,613,496,93	Boolean	Code	STRING	3	Inferred
RECORD_ID	12	489	489	100,000,000,00	Identifier	Identifier	STRING	10	Inferred
SS_NUM	13	489	481	98,364,008,18	US Social Security Num	US Social Security	STRING	11	Inferred
ZIP	14	489	99	20,345,398,77	US Zip	US Zip	STRING	5	Inferred



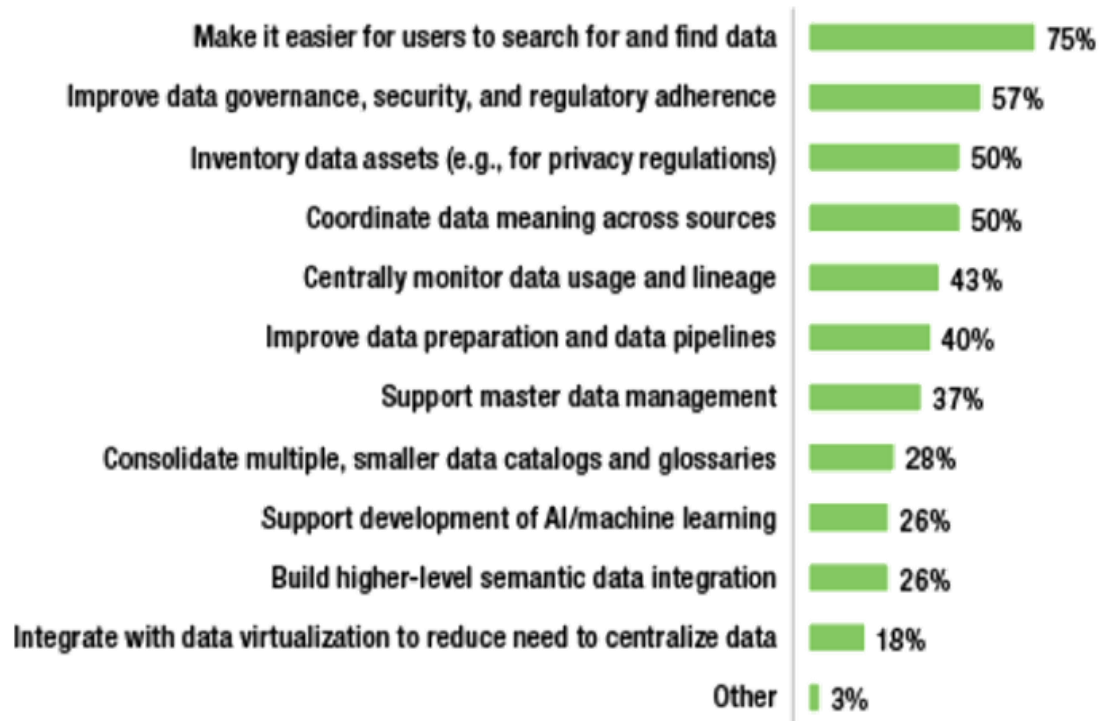
# Unified Platform for Data Integration, Data Quality & Data Governance

## Data Cleansing Approach



# Unified Platform for Data Integration, Data Quality & Data Governance

For your organization, which of the following goals are most important in setting up a centralized data catalog, glossary, or metadata repository?



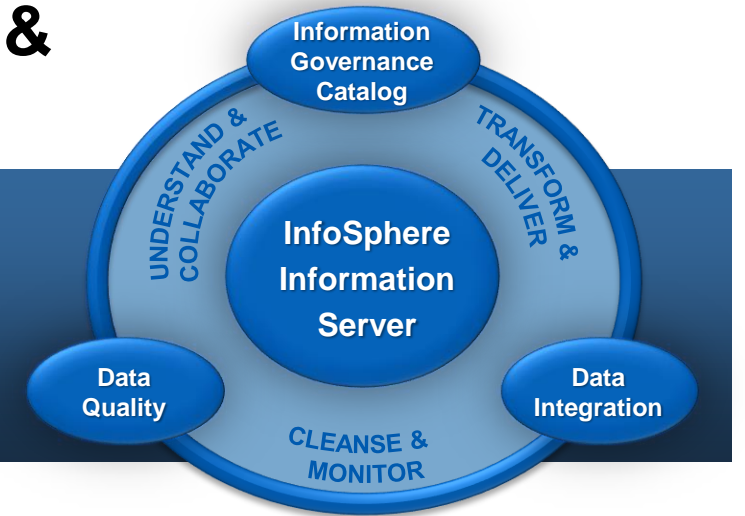
- Making it easier to search for and find data is common goal
- Modern catalogs can supply information at the point of data use
- Valuable when integrated with data virtualization
- 40% say use to improve data prep and pipelines

Source: "Evolving from Traditional BI to Modern Business Analytics," TDWI Best Practices Report, Q3 2020. Based on answers from 183 respondents. Respondents were asked to select their top five goals.

# Unified Platform for Data Integration, Data Quality & Data Governance

## ... Powered by IBM InfoSphere Information Server

Integrating and transforming data and content to deliver accurate, consistent, timely and complete information on a single platform unified by a common metadata layer



### Data Integration

#### *Transform & Deliver*

- Massive scalability
- Power for any complexity
- Deliver in batch and/or real-time with change capture



### Data Quality

#### *Cleanse & Monitor*

- Analyze, classify & validate
- Cleanse & standardize
- Define, manage & monitor data rules & exceptions



### Information Governance Catalog

#### *Understand & Collaborate*

- Catalog technical metadata & align w/ business language
- Manage (big) data lineage
- Manage information governance policies

- common connectivity • shared metadata • common execution engine
- flexible deployment: On-premise, Grid, Cluster, Cloud or native on Hadoop



# Unified Platform for Data Integration, Data Quality & Data Governance

## What is Data Governance ?

Data governance is an emerging, cross-functional management program that treats data as an enterprise asset; a collection of corporate policies, standards, processes, people and technology essential to managing critical data to a set of goals.

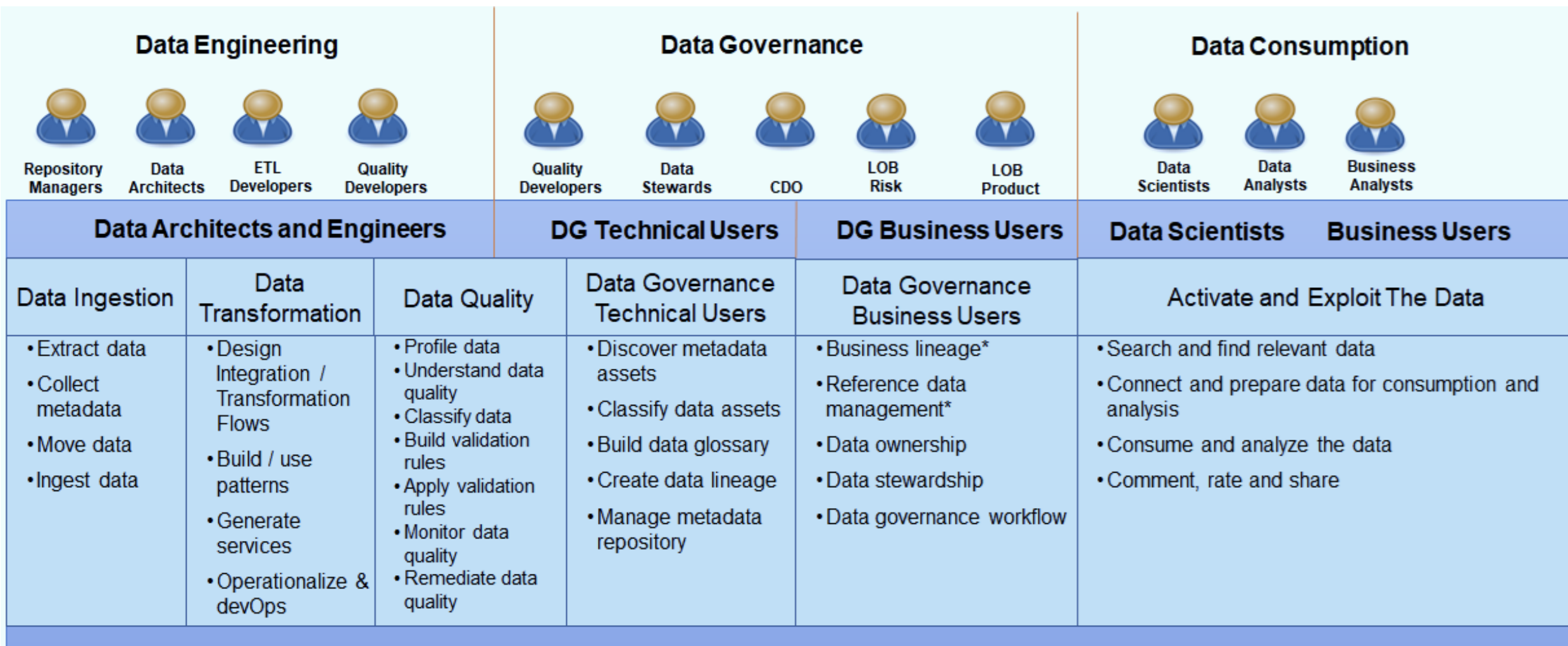
Maria Villar & Theresa Kushner

Data governance is the organization and implementation of policies, procedures, structure, roles, and responsibilities which outline and enforce rules of engagement, decision rights, and accountabilities for the effective management of information assets.

John Ladley & Danette McGilvray

**Objective: To Manage A Valuable Corporate Asset - DATA**

# Unified Platform for Data Integration, Data Quality & Data Governance



# Unified Platform for Data Integration, Data Quality & Data Governance

## InfoSphere Information Server – Information Governance Catalog

Understand and collaborate

**Most comprehensive governance solution for end-to-end visibility to your metadata**

### Understand

- Leverage a comprehensive & rich catalog of information assets
- Provides business context for IT assets
- Dramatically increases business confidence in information assets

### Govern

- Collaboratively establish a governed business vocabulary
- Create stewards, assign responsibilities
- Rely on end-to-end lineage across information assets
- Links business terms & information governance rules to information assets & operational rules
- Open interface to create, import and manage extensions for Data Sources and Data Flows

Welcome to InfoSphere Information Governance Catalog

IBM InfoSphere Information Governance Catalog enables the enterprise to expose a central catalog of glossary assets and information assets to end users. The meaning of those assets can be broadened by adding technical descriptors and business context.

InfoSphere Information Governance Catalog provides search, browse, and query capabilities. In addition, you can establish asset collections and run lineage reports to examine data flow between assets.

Search

Search for Terms [Search] [Options]

InfoSphere Information Governance Catalog

Database Table Details

PROD\_MRT  
Production Mart for product transaction reporting.

Context: [Database] [Schema] [Table]

Stewards (1): [Mr. Marc Haber]

Assigned to Terms (1): [Product] [Retail Glossary] [Business Concepts] [Product]

Asset Name	Type	Length	Primary Key	Position	Allow Null Values	Unique	Data Rules
ID	INT32	0	False	0	False	False	
ITEM_NAME	STRING	100	False	0	False	False	
ITEM_NUM	INT32	0	False	0	False	False	
ITEM_UNITS	INT32	0	False	0	False	False	
PLANT	STRING	100	False	0	False	False	

Database Table Usage Information (1)

Written by (Design) (1): [PROD\_MRT] [MARC-203] [EWS] [EWS\_ProductMart]

Database Table Lineage

Database Table Lineage diagram showing data flow from various sources (e.g., PROD\_MRT, PROD\_MRT\_2, PROD\_MRT\_3) to the central PROD\_MRT table.



# Unified Platform for Data Integration, Data Quality & **Data Governance**

## **Data Governance** — **IBM Infosphere**

IBM InfoSphere Information Governance Catalog (IGC) allows companies to

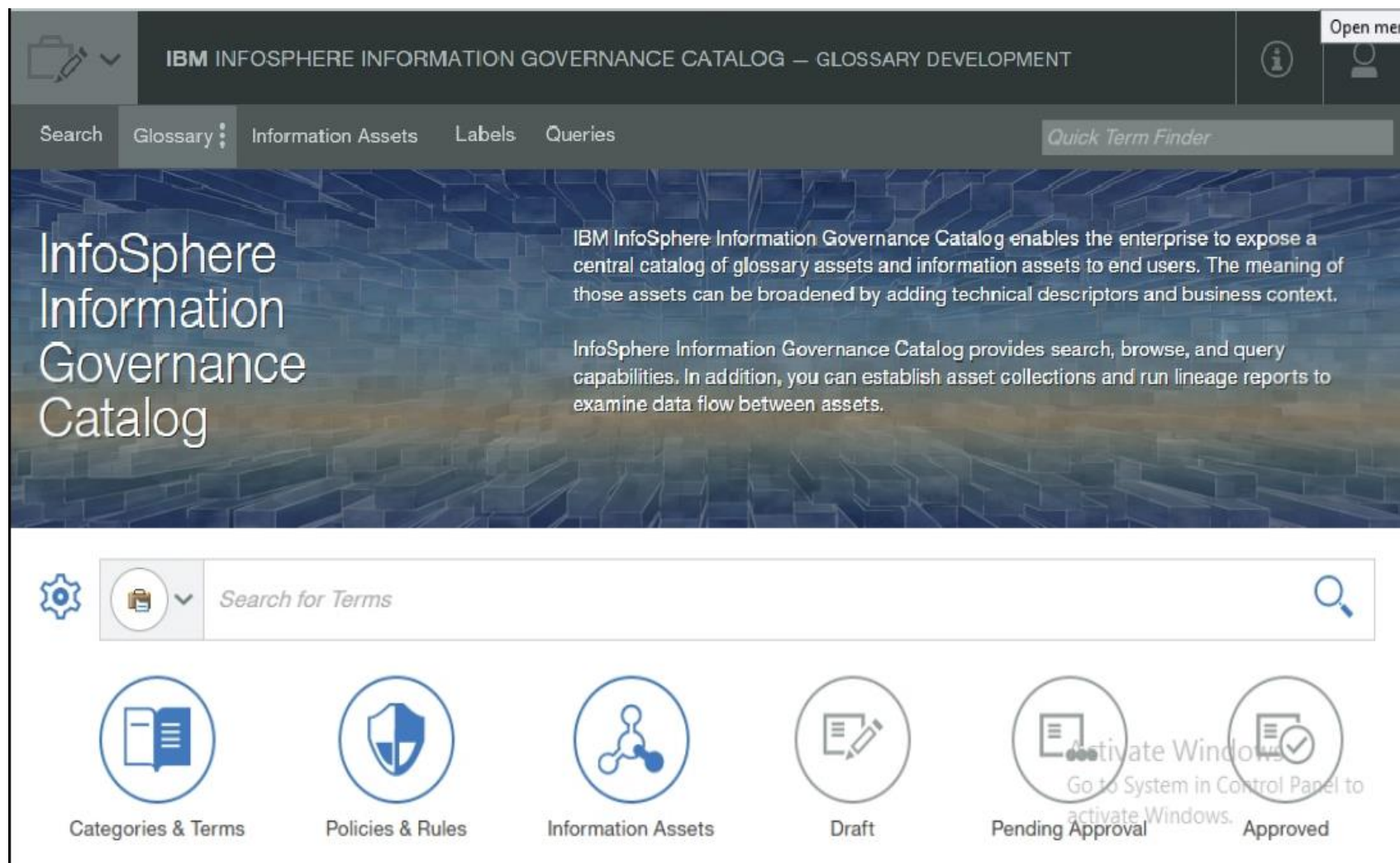
- govern their data and
- data movement processes by
  - exposing, and relating, all types of data and applications, to the business processes, rules, and initiatives, that use that data.
  - allows analysts and consumers to find and examine assets based on their usage and business purpose, without technical skills and knowledge of back-end systems.

Key capabilities:

- Establish a centralized common Business Glossary
- Apply Governance Policies and Rules
- Define Business Processes and Applications using Custom Attributes
- Analyze Data and Business Lineage Graphs and Reports

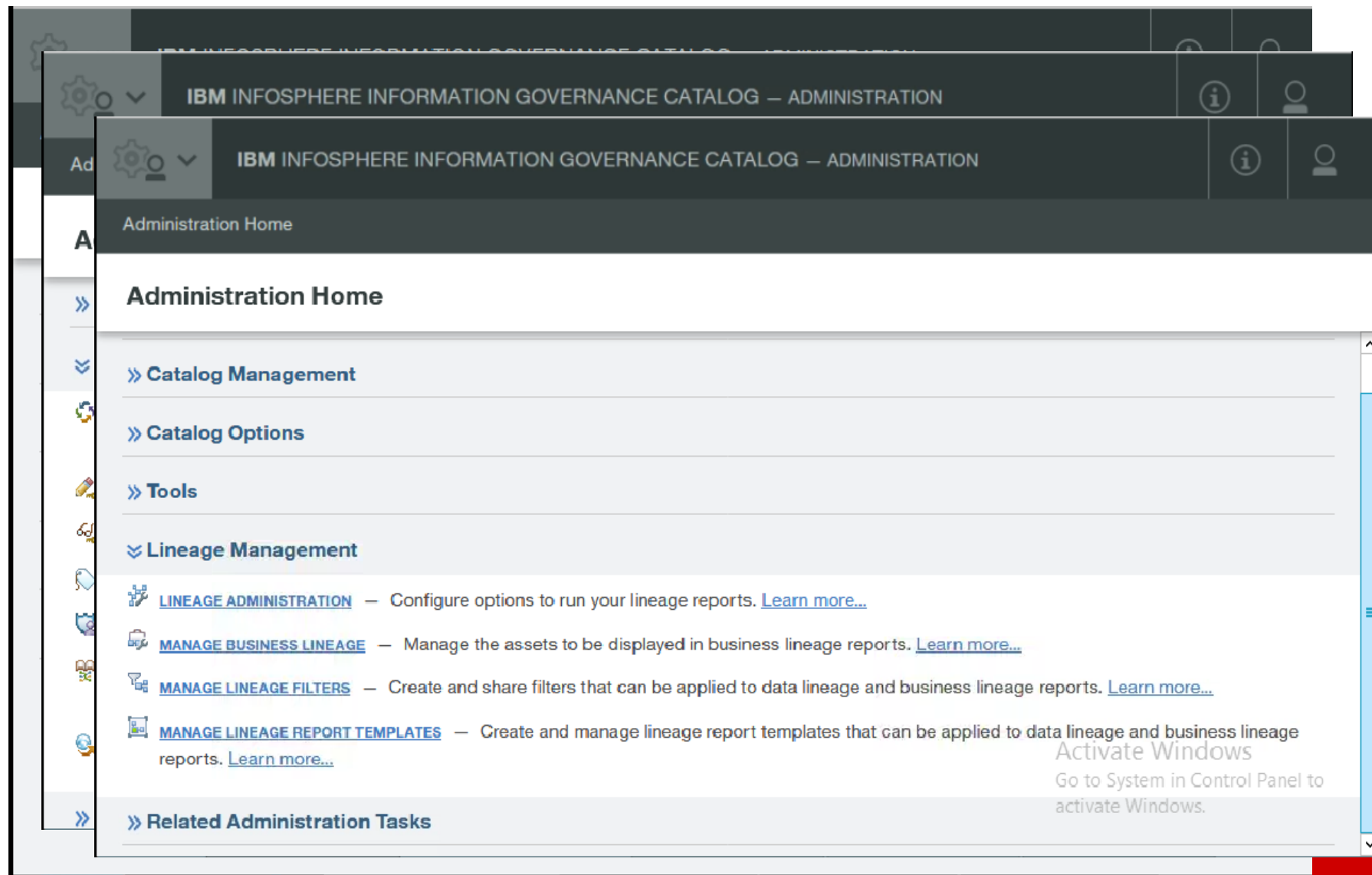
# Unified Platform for Data Integration, Data Quality & Data Governance

## Data Governance — IBM Infosphere



# Unified Platform for Data Integration, Data Quality & Data Governance

## Data Governance — IBM Infosphere



# Unified Platform for Data Integration, Data Quality & Data Governance

## Data Governance – IBM Infosphere

