# Introducing Data Quality into the Information Supply Chain

Sue Eris – Technical Sales

01/05/2012

# What we'll discuss

- Overview of Data Quality in the Information Supply Chain
- Challenges of Poor Data Quality
- Ensuring Data Quality Across the Enterprise
    - Understand & Define
    - Cleanse & Standardise
    - Monitor & Govern continuously
- Demo of InfoSphere Information Server

twitter: *Follow @ANZ_IM or mention #IIGS*

# Too many organisations are operating with blind spots

## 1 in 3
Business leaders **frequently make decisions** based on information they don't trust, or don't have

## $5.7M USD
Lost time **reformatting information** between applications

## 1 in 2
Business leaders say **they don't have access to the information they need** to do their jobs

## $5.3M USD
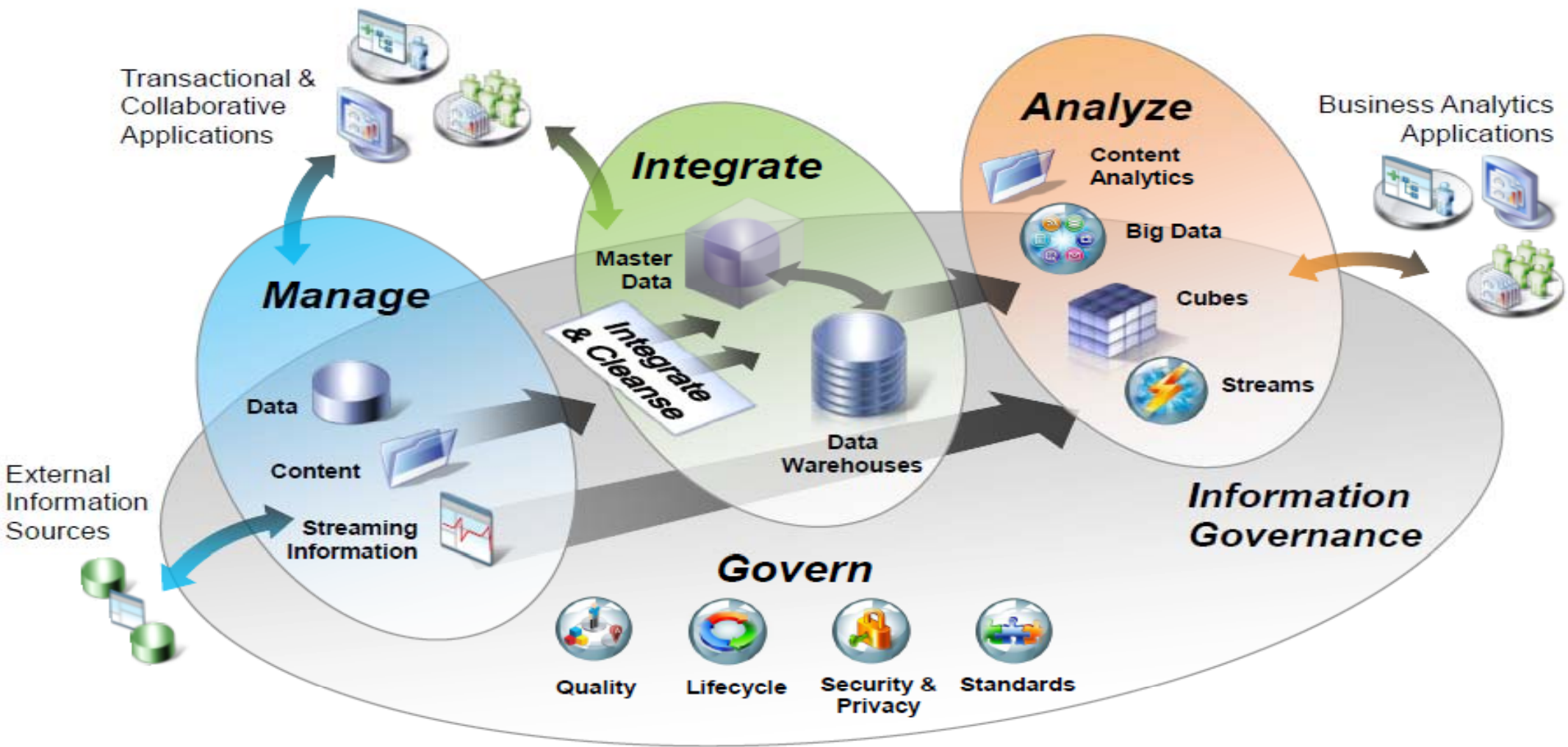Lost time **searching for the right information**

*Annual cost per 1,000 knowledge workers*

Sources: IBM: Break Away with Business Analytics and Optimization Study; IDC

3

twitter: *Follow @ANZ_IM or mention #IIGS*

# Success requires governance across the "Information Supply Chain"



Transactional & Collaborative Applications

External Information Sources

**Manage**

Data

Content

Streaming Information

**Integrate**

Master Data

Integrate & Cleanse

Data Warehouses

**Analyze**

Content Analytics

Big Data

Cubes

Streams

Business Analytics Applications

**Govern**

Quality

Lifecycle

Security & Privacy

Standards

**Information Governance**

◆ **Trusted** ◆ **Relevant** ◆ **Governed**

twitter: *Follow @ANZ_IM or mention #IIGS*

4

# Data Quality in the Information Supply Chain

- **Define** what 'high quality' data means within your business
- **Standardize and cleanse** data to enable strategic initiatives
- **Constantly monitor** for data quality issues and understand how to deal with an exception when it occurs
- Understand the **lineage** of your data



Quality Management

Monitor & Govern

Understand & Define

Cleanse & Standardize
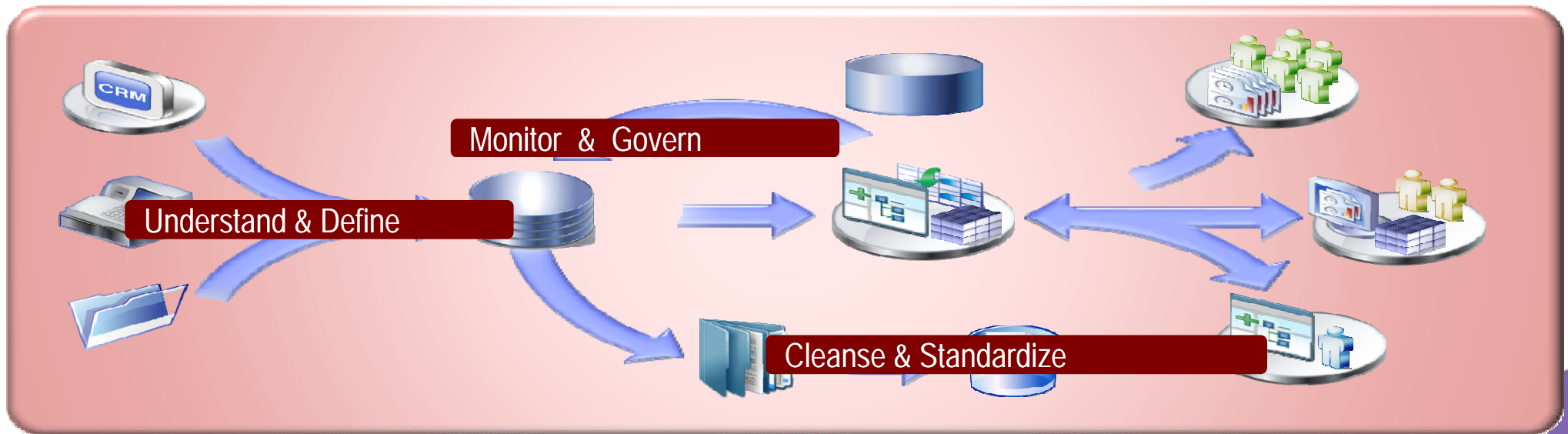
# Organisations continue to have Data Quality Challenges

- *Compliance and transparency pressures increasingly highlight data quality issues*
  - No method to maintain high quality data

- *Unreliable insights*
  - Low data quality leads to lack of trust and results in poor business decisions
  - Inability to identify source of quality issues
  - Unreliable insights are persisted to other strategic initiatives, which then base key business decisions on bad data

- *High costs & negative customer satisfaction*
  - Organisations are recognizing that there can be both direct (missed revenue opportunity) and indirect (low customer satisfaction and high churn) financial costs from poor data quality.

twitter: *Follow @ANZ_IM or mention #IIGS*

# Can't Define Quality If You Don't Understand Your Data

**Understand & Define**

**Challenge:**

- Distributed heterogeneous sources
- No documentation on data structures and data relationships
- Inconsistent data definitions between Business User and Technical User
- Lack of trusted data – unknown quality
- Limited understanding of confidential data elements
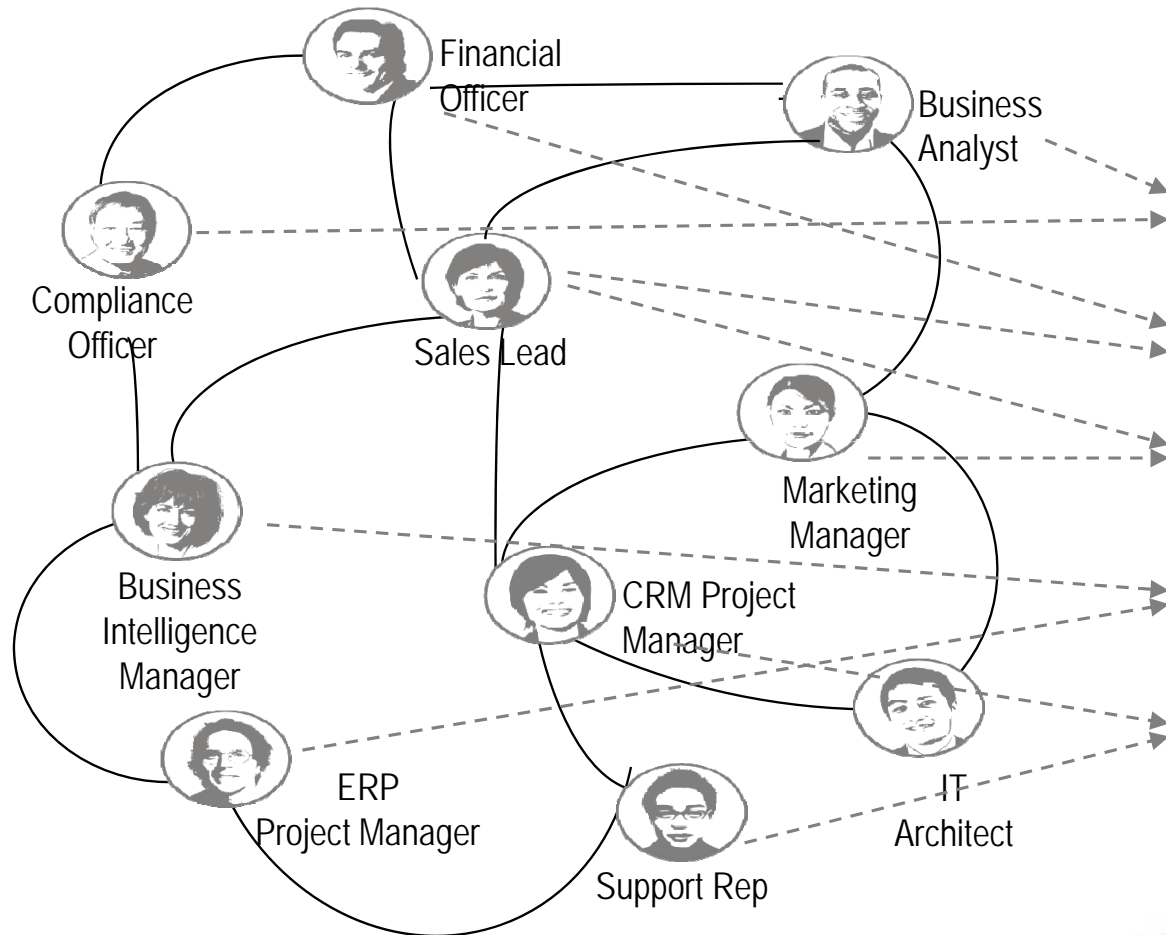- Data does not stand still, new data always added, but not necessarily evaluated for quality

**Cost Prohibitive Alternative Solutions:**

- Manual spot checking of data not feasible or reliable

twitter: *Follow @ANZ_IM or mention #IIGS*

# Lack of Consistent Terminology Propagates Poor Data Quality

IBM



Financial Officer

Business Analyst

Compliance Officer

Sales Lead

Business Intelligence Manager

Marketing Manager

CRM Project Manager

ERP Project Manager

IT Architect

Support Rep

*How does each user define:*

*"Active Subscriber"?*

❑ *Mobile user who has used "any" service in the mobile network*

❑ *User who paid for the service at least 1 time in the past 90 days.*

❑ *Mobile user who has a phone plan, but not SMS*

❑ *Only post-paid customers, not pre-paid customers*

❑ *User who makes at least 1 call over the period of 90 days*

twitter: *Follow @ANZ_IM or mention #IIGS*

8

# Why Should I Care About Cleansing Information

Lack of information standards

Data surprises in individual fields

Information buried in free-form fields

Data myopia

The redundancy nightmare

9

| Kate A. Roberts | 416 St Kilda Rd #2, | | Saint Kilda, | VIC | 3182 |

| Catherine Roberts | Four sixteen St Kilda APT2, | | Elwood, | VICT | 3184 |

| Mrs. K. Roberts | 416 Saint Kilda Unit 2, | | St Kilda | | 3100 |

| Name | TFN | Telephone |
| --- | --- | --- |
| J Smith DBA Lime Cons. | 28-02-1975 | 0391733803 |
| Williams & Co. C/O Bill | 025-37-1888 | 415-392-2000 |
| 1st Natl Provident | 34-2671433 | 3380321 EXT 99 |
| HP 15 State St. | 508-466-1200 | Fax number |

| Name | Notes |
| --- | --- |
| Customer A | Customer is not to be called after 8pm |
| Customer B | There is outstanding issue, priority 2A. Required fix by 01/02 |
| Customer C | Recent mail have been returned, new addr avail? |
| Customer D | High value customer for special attn |

WING ASSY DRILL 4 HOLE USE 5J868A HEXBOLT 1/4 INCH

WING ASSEMBY, USE 5J868-A HEX BOLT .25" - DRILL FOUR HOLES

USE 4 5J868A BOLTS (HEX .25) - DRILL HOLES FOR EA ON WING ASSEM

RUDDER, TAP 6 WHOLES, SECURE W/KL2301 RIVETS (10 CM)

| 90328574 | IBM | 60 City Road Southgate VIC 3006 |
| 90328575 | I.B.M. Inc. | 60 City Road Southgate VIC 3006 |
| 90238495 | Int. Bus. Machines | 60 City Road Southgate VIC 3006 |
| 90233479 | International Bus. M. | 60 City Road Southgate VIC 3006 |

# Monitor Data Quality with Integrated Data Rules

❑   Data rules enable data quality monitoring over time and leverage data analysis at different points during an integration project

❑   Helps establish benchmarks and baselines to help track data quality – is it deteriorating or remaining constant?

❑   Is used to place indicators/flags in enterprise data which can be tested for valid/invalid values
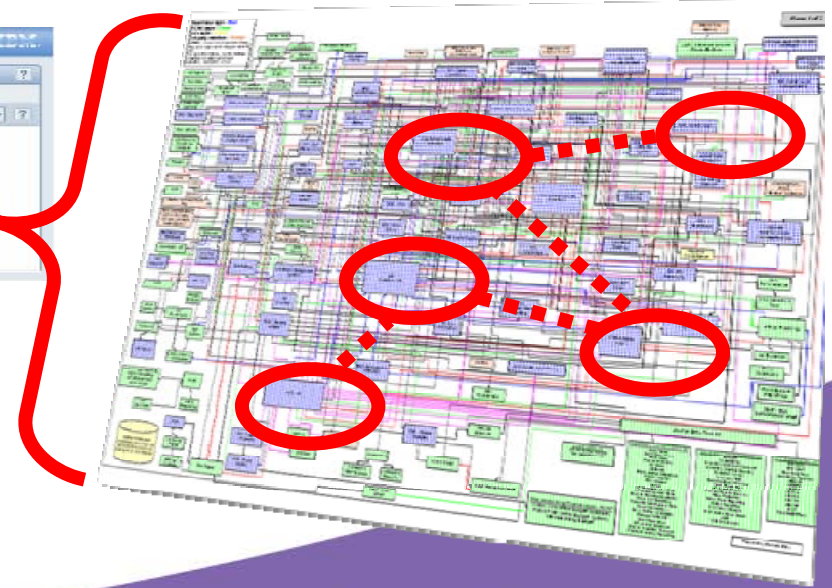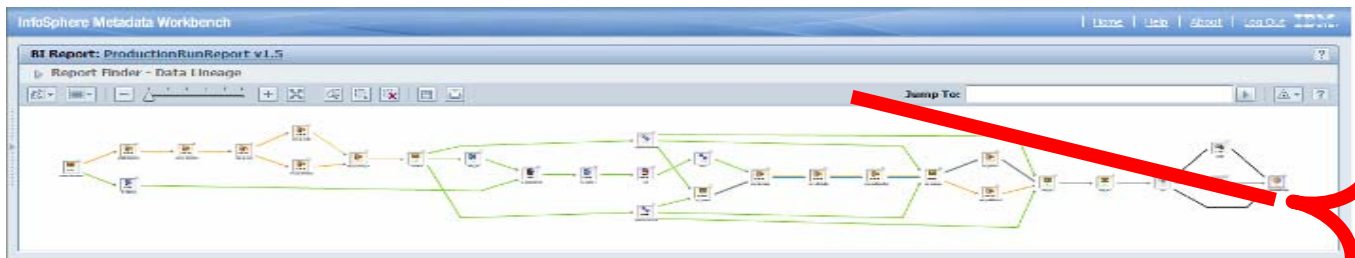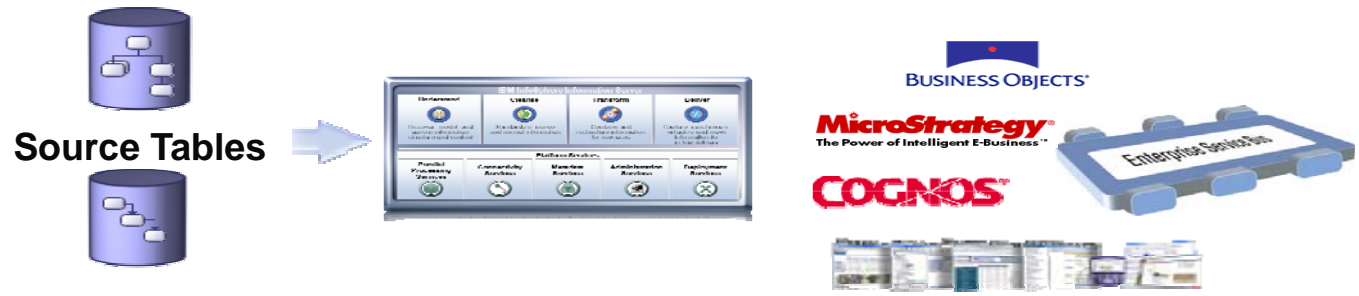
❑   Flag bad data for audit

twitter: *Follow @ANZ_IM or mention #IIGS*

# Govern the Information Supply Chain

**Source Tables**

**Monitor & Govern**

BUSINESS OBJECTS

MicroStrategy
The Power of Intelligent E-Business

COGNOS

Enterprise Service Bus

InfoSphere Metadata Workbench

- Understand the impact of change
- Trace data lineage from sources, through processing and all the way through to data consumption

twitter: *Follow @ANZ_IM or mention #IIGS*

# IBM Provides Solutions for Creating High Quality Data

- A comprehensive Information Governance strategy for creating high quality data that helps you…
  - Ensure that your data contains accurate and complete information you can trust; while reducing the risk of propagating bad information throughout your organization.
  - Build a common understanding of your data across the enterprise among business and technical users.
  - Establish defined data quality rules or standards and continuously monitor the flow of data through the information lifecycle to enforce adherence to those rules and standards.
  - Provide the means to rapidly investigate and resolve data quality issues when they do occur.

**Understand & Define**  **Cleanse & Standardize**  **Monitor & Govern**

twitter: *Follow @ANZ_IM or mention #IIGS*

# IBM InfoSphere Information Server

## *A Platform for Data Quality and Information Governance*

- Provides the framework for defining a business vocabulary
- Promotes Stewardship
- Measures and monitors data quality
- Encourages collaboration
- "Touches" *everything*
  - Source and target data throughout the enterprise
  - Data movement and transformation
  - Business intelligence
- Illustrates end-to-end data lineage
- Powerful functionality surrounding a common "core" of managed metadata



twitter: *Follow @ANZ_IM or mention #IIGS*
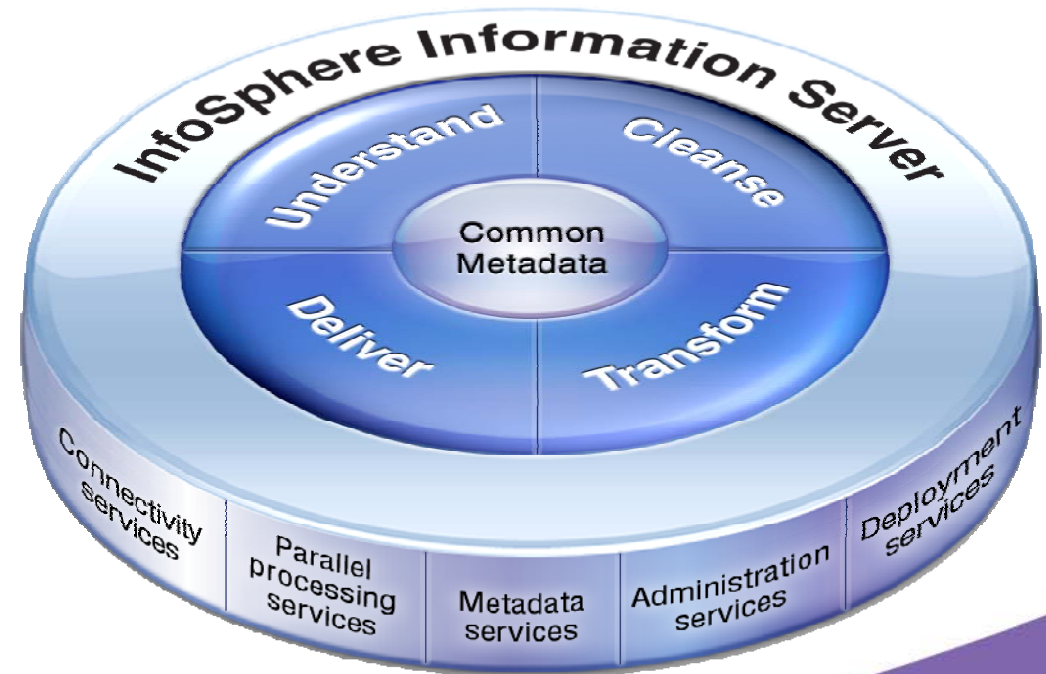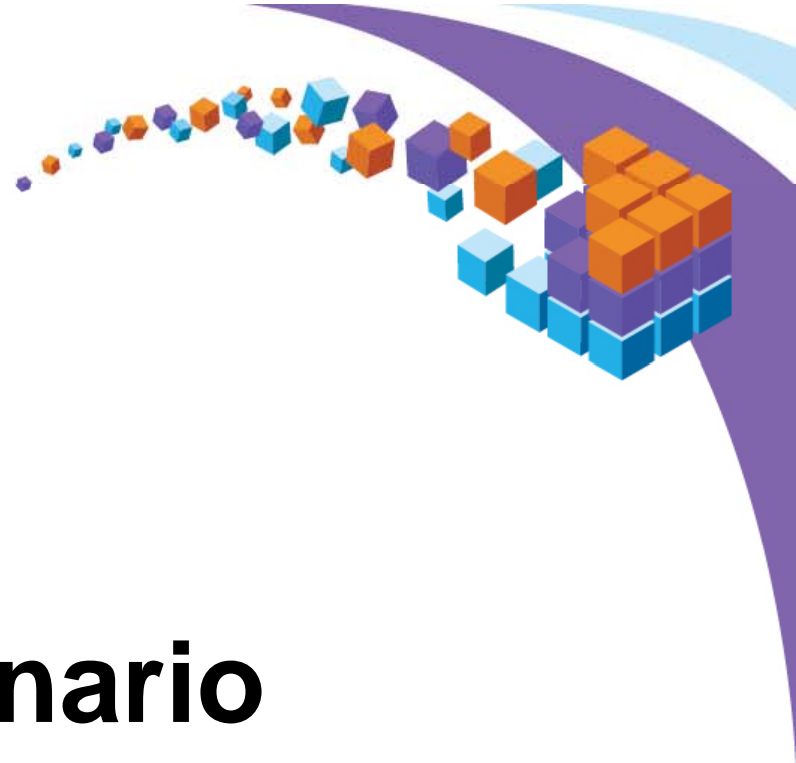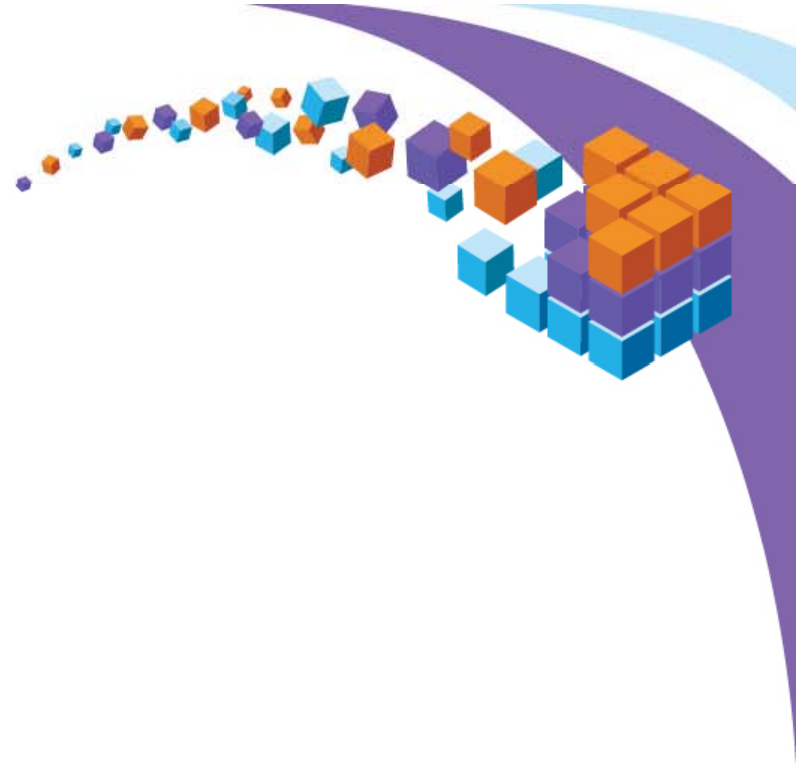
# Demonstration Scenario

# Business Problem

- Company had multiple sources
- Need to integrate and govern, but...
  - Do not know what they do not know
  - Do not know where the critical data is
  - People that know the sources are gone
  - Heterogeneous sources

# Business ~~Problem~~ Solution
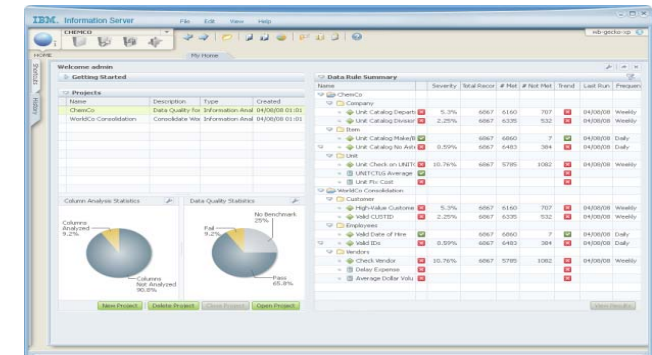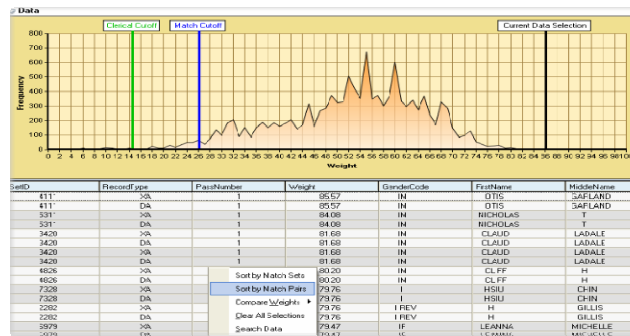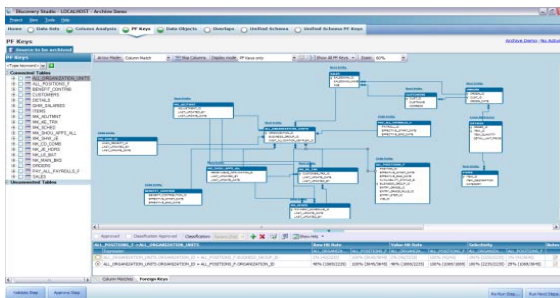
**Understand**

**Cleanse**

**Monitor**

16

# Demonstration

# InfoSphere – the only vendor who is a leader in every category of Information Integration and Governance

| | | |
|---|---|---|
| **Leader!** | **Information Integration & Data Quality** | **27.5%** market share |
| **Leader!** | **Master Data Management** | **40%** market share |
| **Leader!** | **Information Lifecycle Management** | **76%** market share |
| **Leader!** | **Privacy & Security** | **"IBM InfoSphere Guardium has been deployed across many large enterprises"** |

twitter: Follow @ANZ_IM or mention #IIGS

twitter: *Follow @ANZ_IM or mention #IIGS*