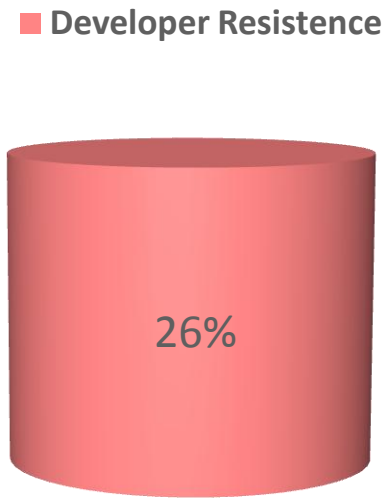
Oracle Data Integration & Governance Workshop

ROI Of Data Governance

Milomir Vojvodic
Director Business Development - Data Integration EMEA

Highest Friction Point To DG&DI Adoption



Source : Info-Tech Group “Friction Points To DI Adoption”

Data Integration Competency Center

And

Total Economical Impact

Integration Competency Center

Data Integration

•Reflects the convergence of Enterprise Application Integration (EAI), Data Integration (DI) and Extract, Transform, load (ETL) vendors.
•Encompasses data consolidation, federation, propagation and access.
•Is gaining momentum as enterprises move from hand-coding to tools.
•Prevents silos of information and enables an enterprise-wide view of data.

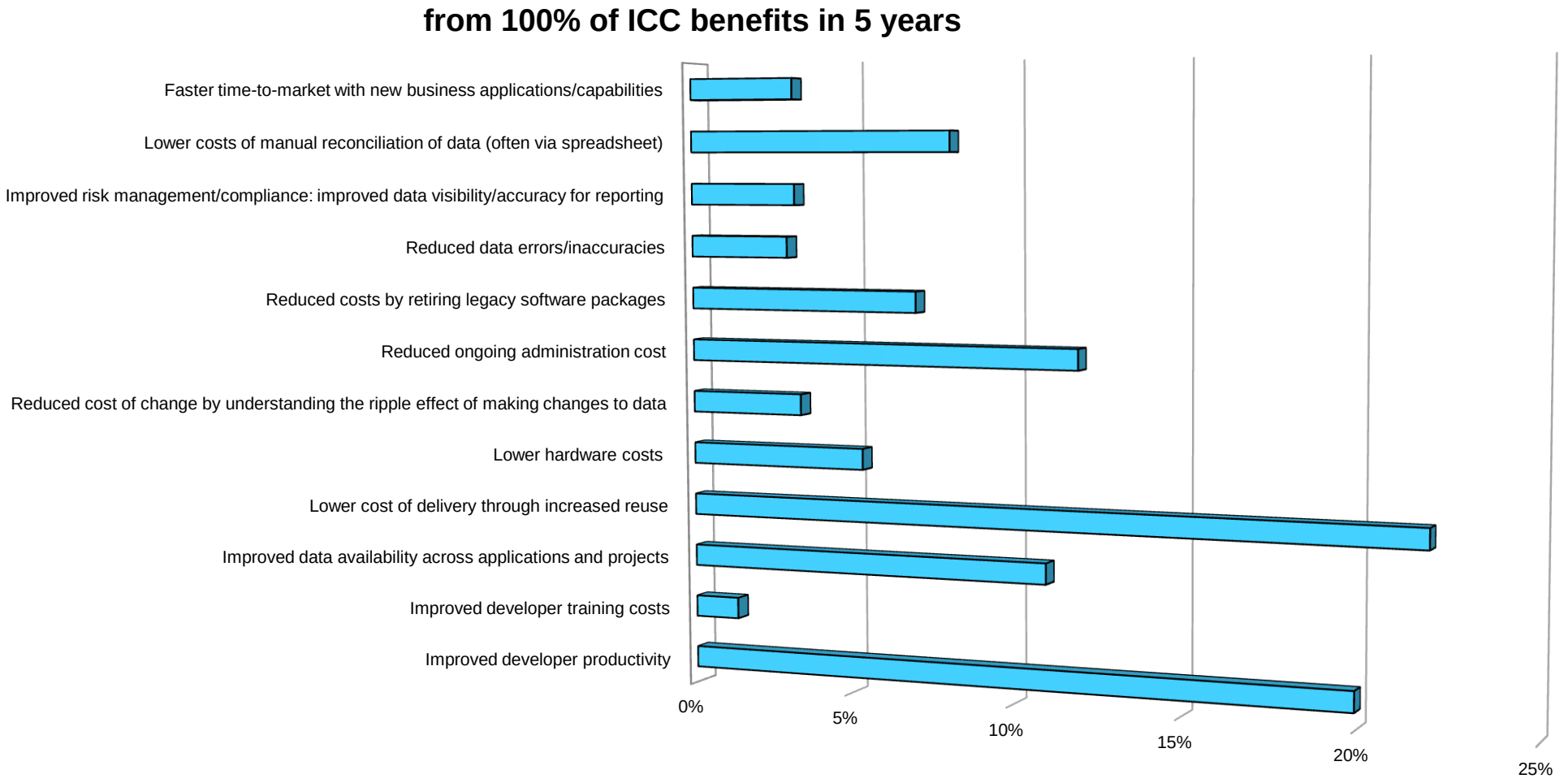
Process Integration

•Is workflow automation for processes, e.g. online ordering.
•Emphasizes a top-down process design approach to a greater extent than the other two, starting from a process model.
•Allows organizations to streamline process and creates opportunity for coordinated escalation

SOA & Middleware

•Is used to connect enterprise apps without transforming the data.
•Creates access to data in distributed architectures.
•Requires advanced architecture to achieve success.

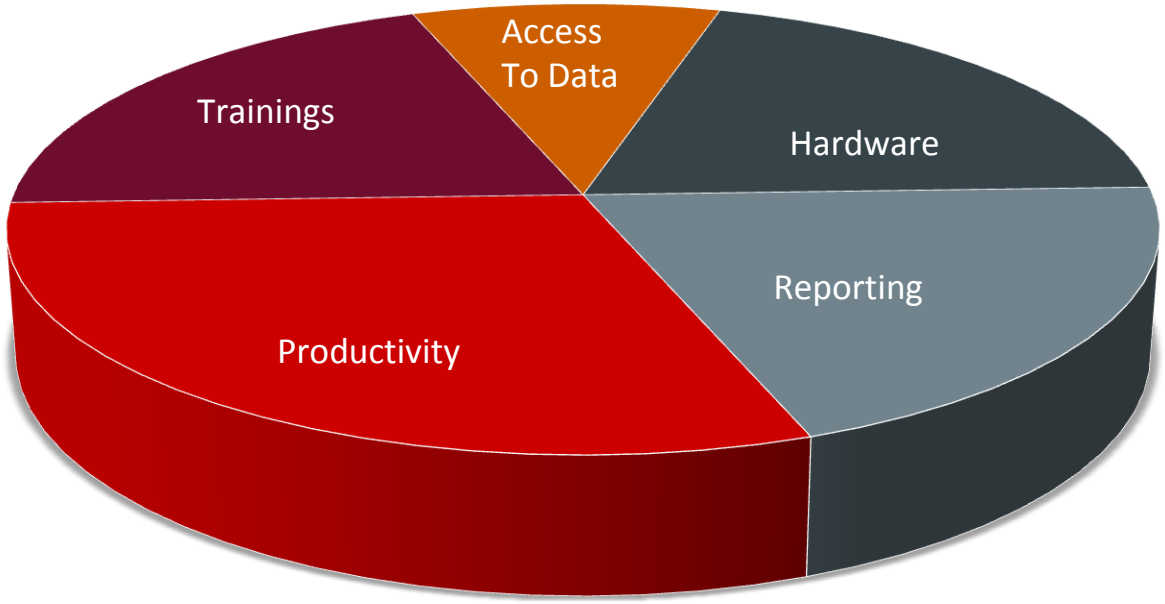
Distribution Of Savings In 5 Years



Source : INFA

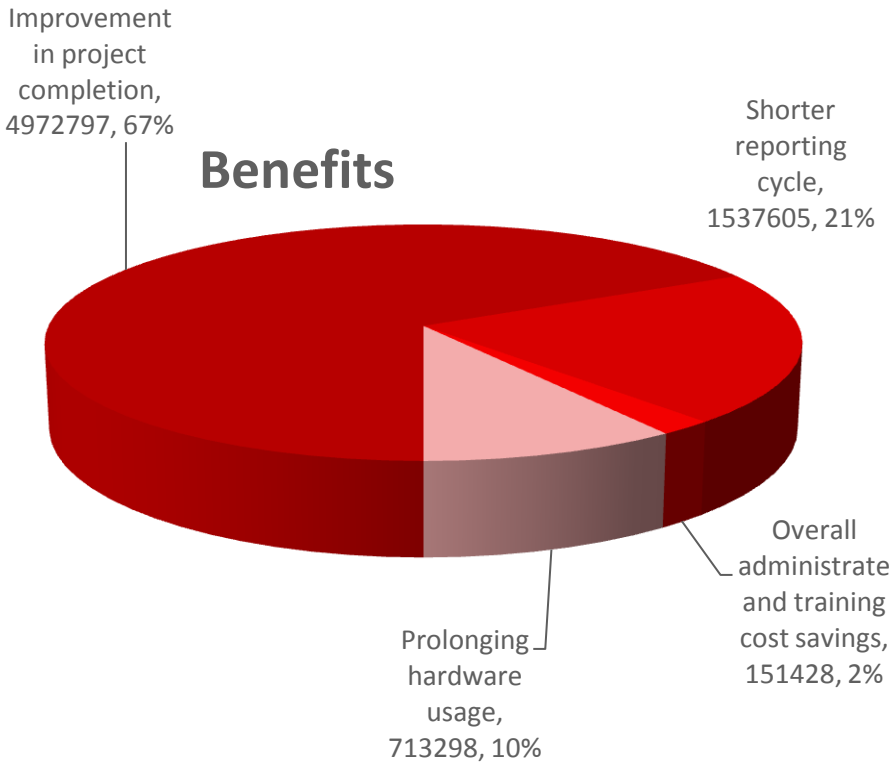
Distribution Of Savings In 5 Years

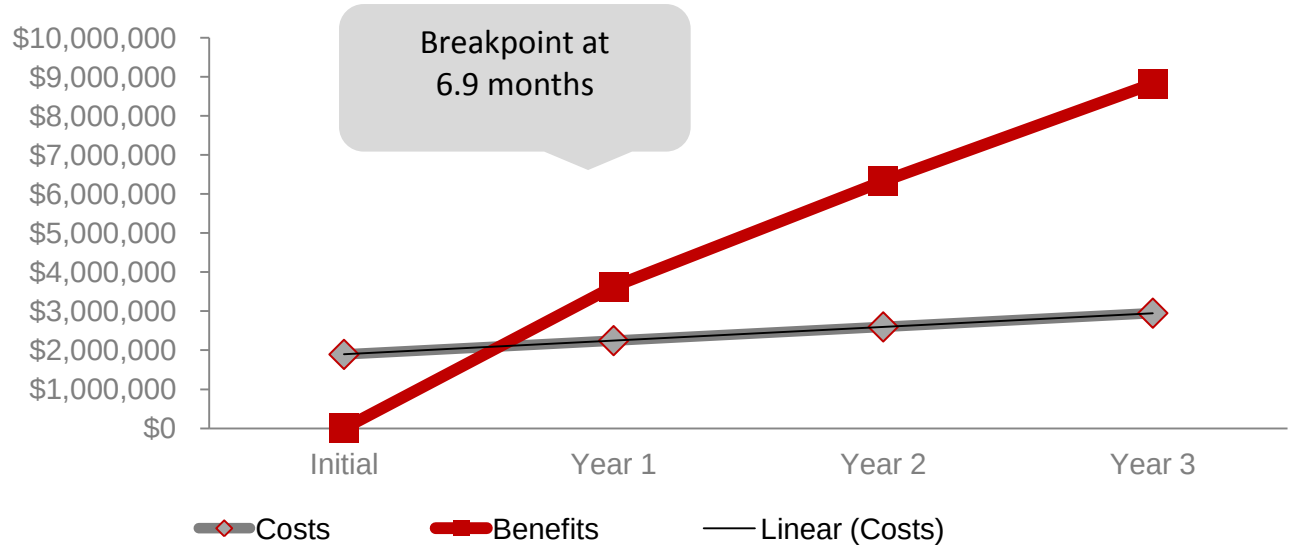
from 100% of Data Integration
CC benefits in 3 years

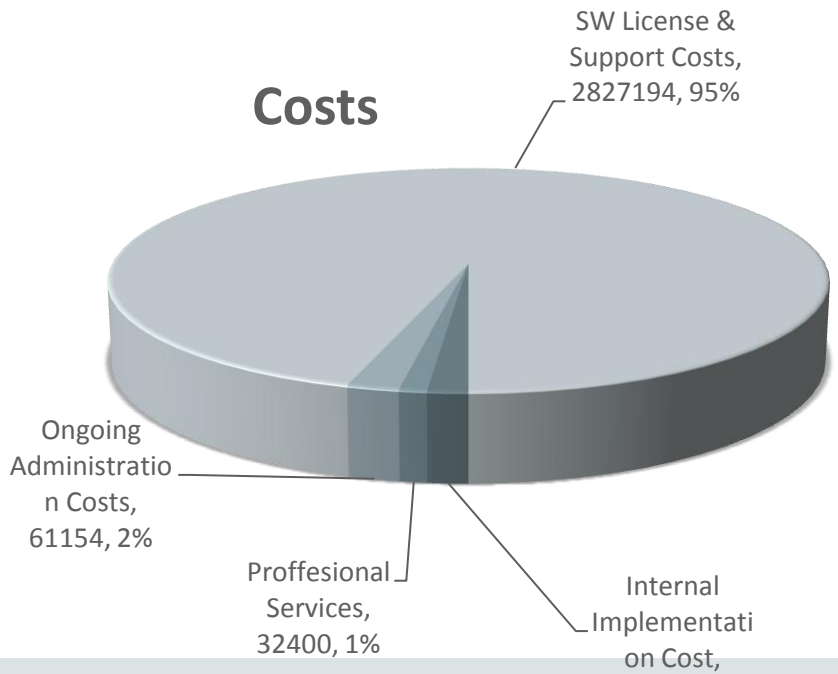


Source : Analysis Of Oracle Customer References

ICC Life Cycle







Analytical & Operational Data Integration

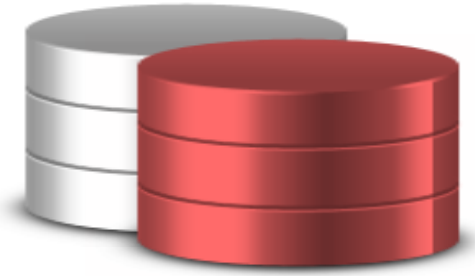
Oracle GoldenGate
Real Time Replica Between Any Databases

Active Data Guard
Disaster Recovery Of Oracle DB EE

Oracle Data Integrator
ETL Platform

Enterprise Data Quality
Data Quality Platform

Enterprise Metadata Manager
Data Quality Platform



OGG → OEMM, ODI, OEDQ → Zero Downtime Migrations & Upgrades → New DB/HW/OS/APP

Migrations & Consolidations

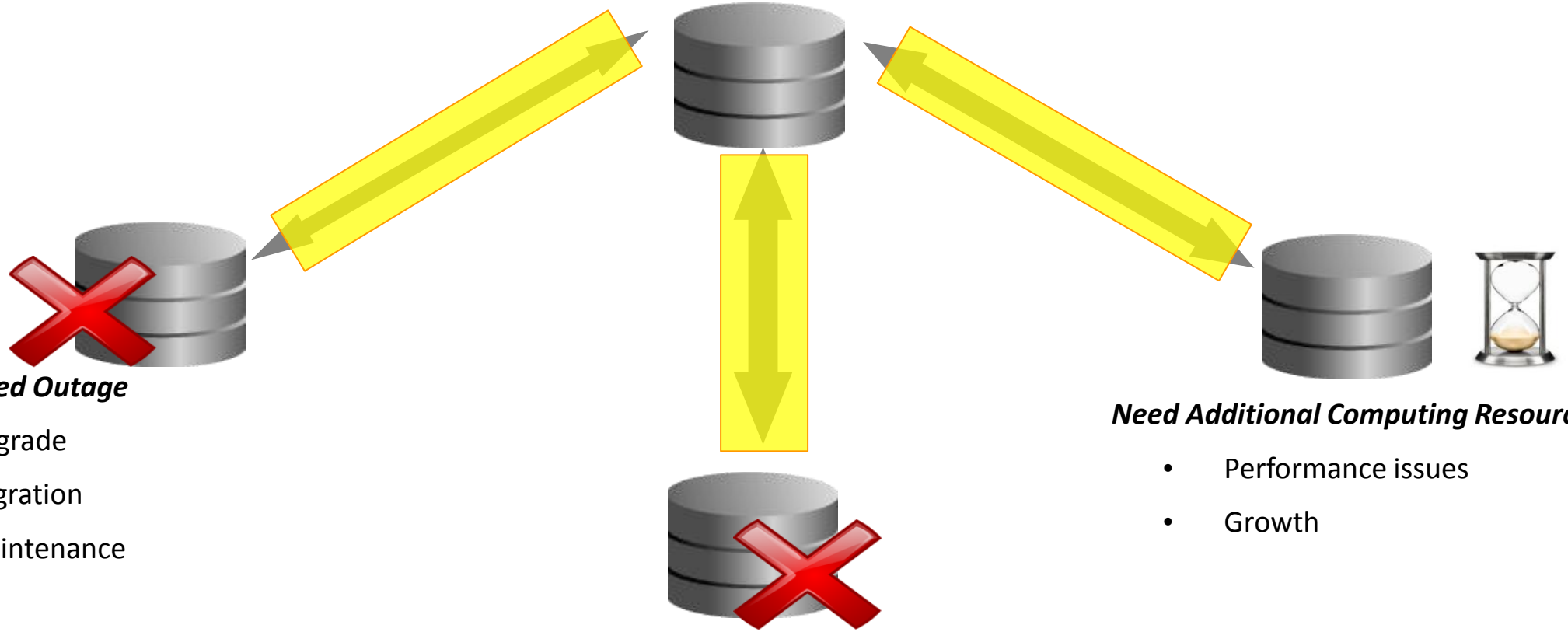
OGG → Active/Active High Availability → Fully Active DB

OGG + ADG → Query Off-Loading and Disaster Recovery → Reporting Database and/or DR database

OGG → OEMM, ODI, OEDQ → BI&DW Synchronization and Loading → Data Warehouse

Reusability Of DIS Licenses

Product Needs To Be Available – No Issues



Need Planned Outage

- Upgrade
- Migration
- Maintenance

Need Additional Computing Resources

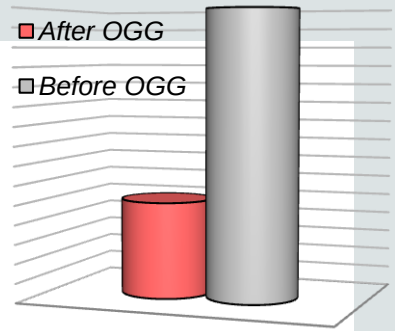
- Performance issues
- Growth

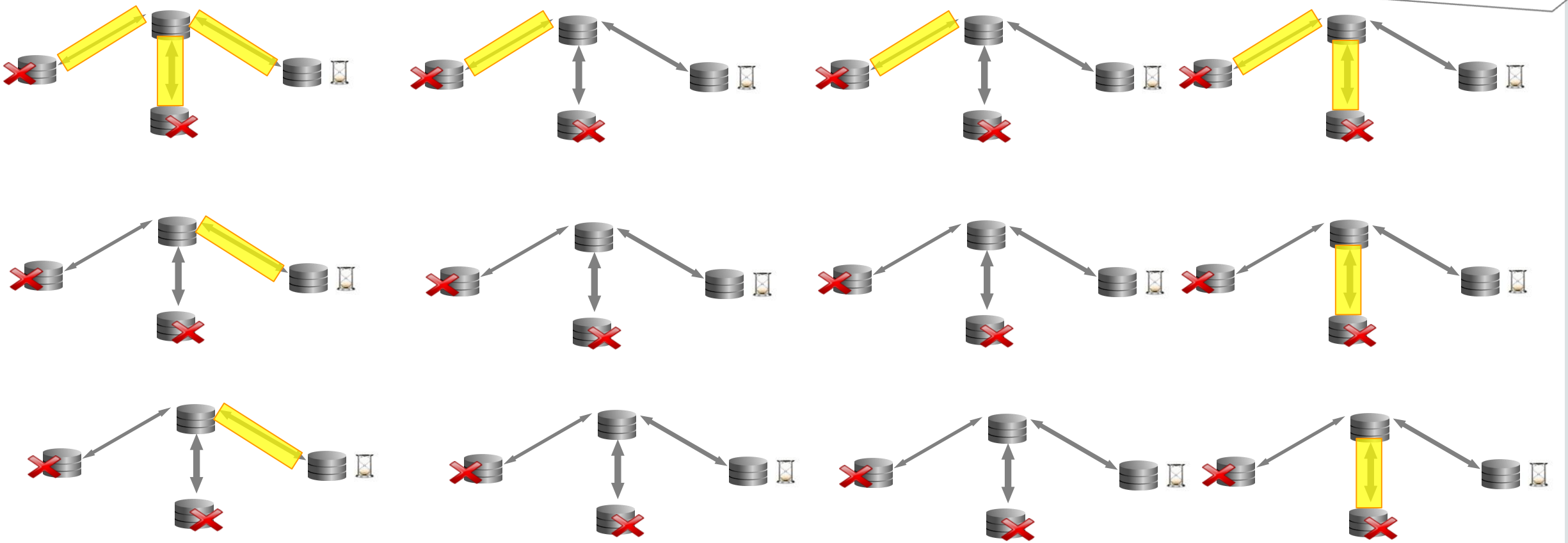
Protect From Unplanned Outage

- System Failure
- Data Failure

Reusability Of DIS Licenses

Communication cost savings and no need for expensive network infrastructure upgrades by 60%





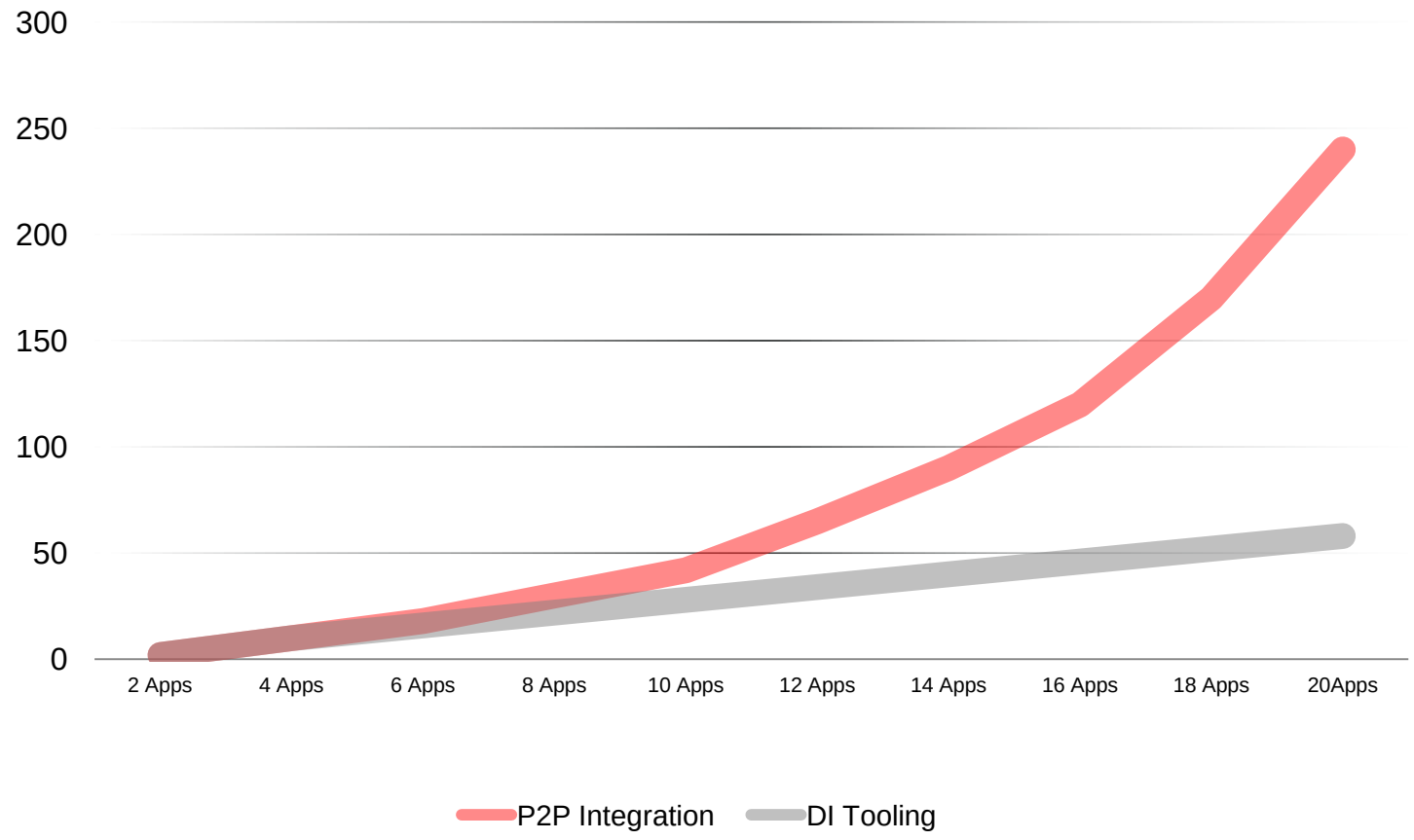
Data Integration Trends

Exponential Chaos

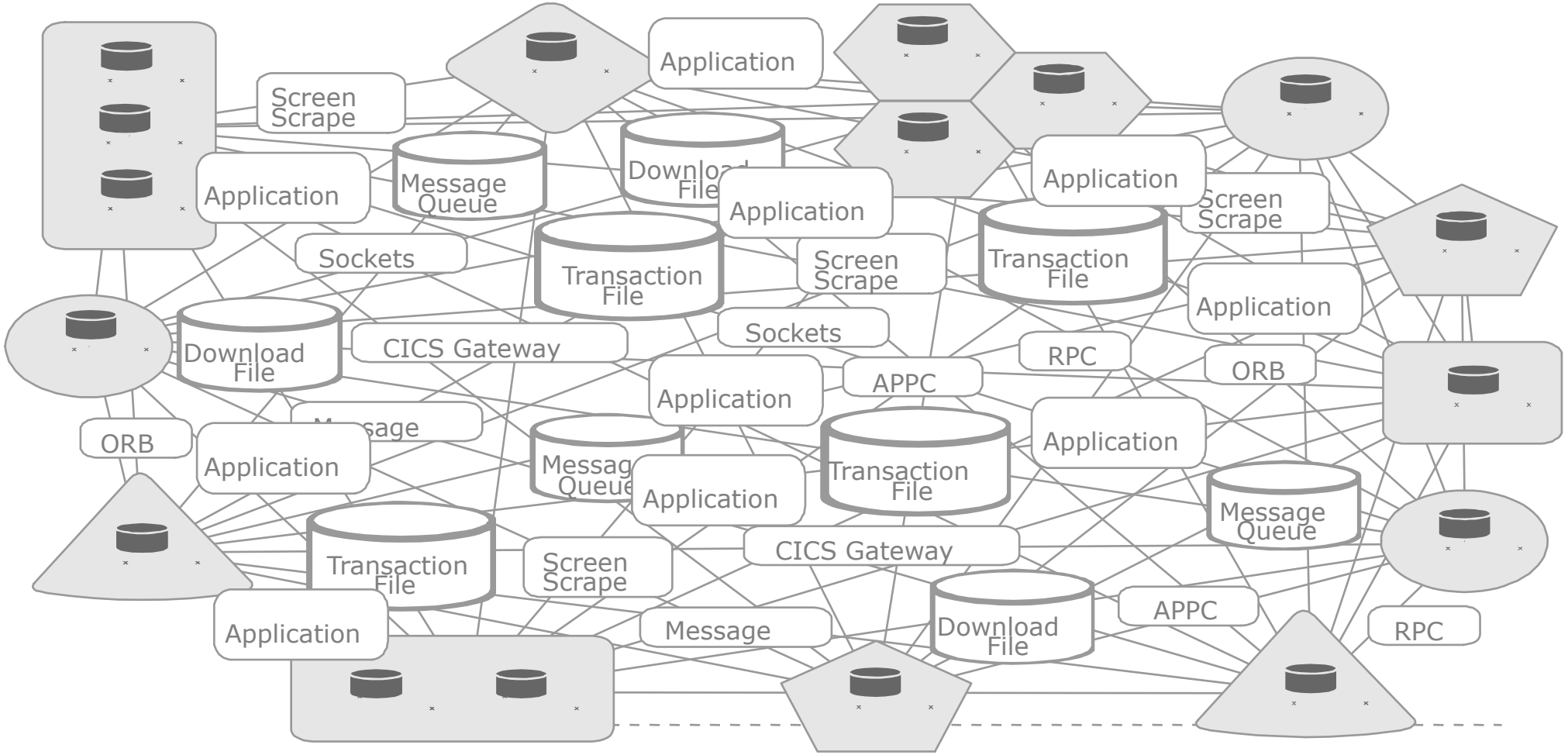
of integration scenarios

Point-to-point integration architectures cannot keep up with business growth

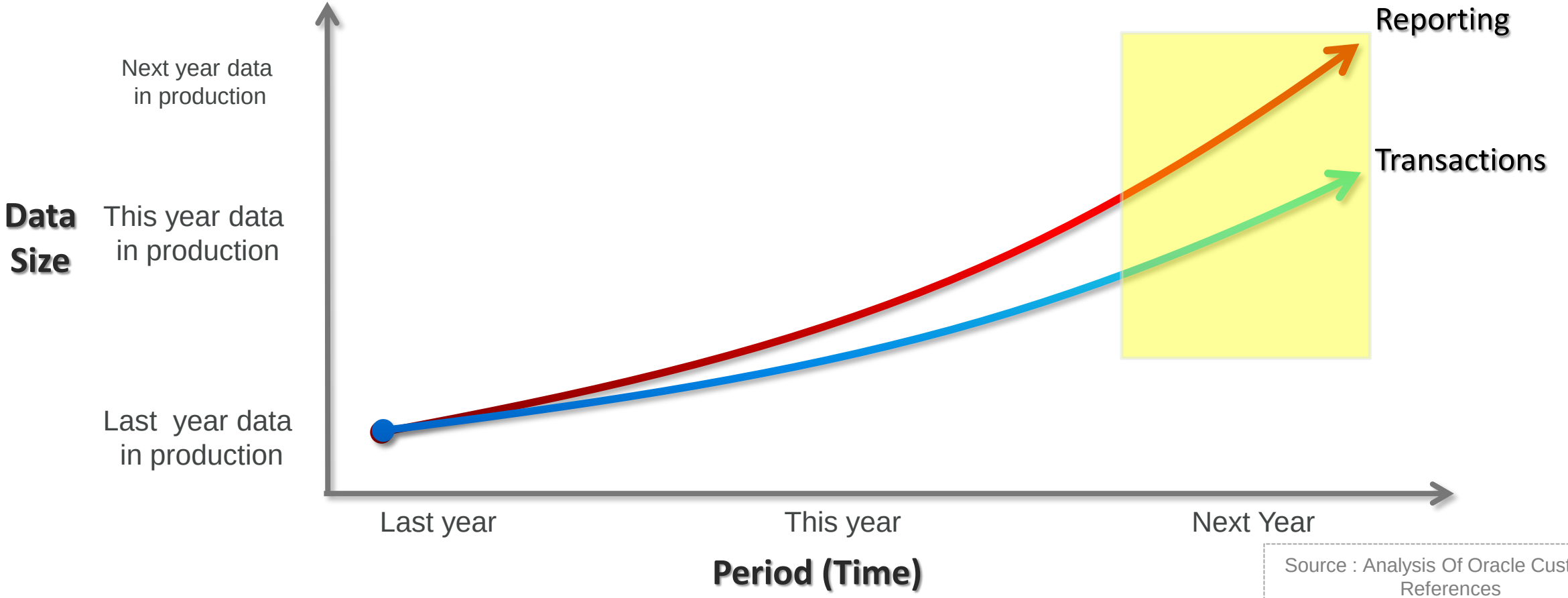
and represent weak links in mission critical integration scenarios.



Architecture

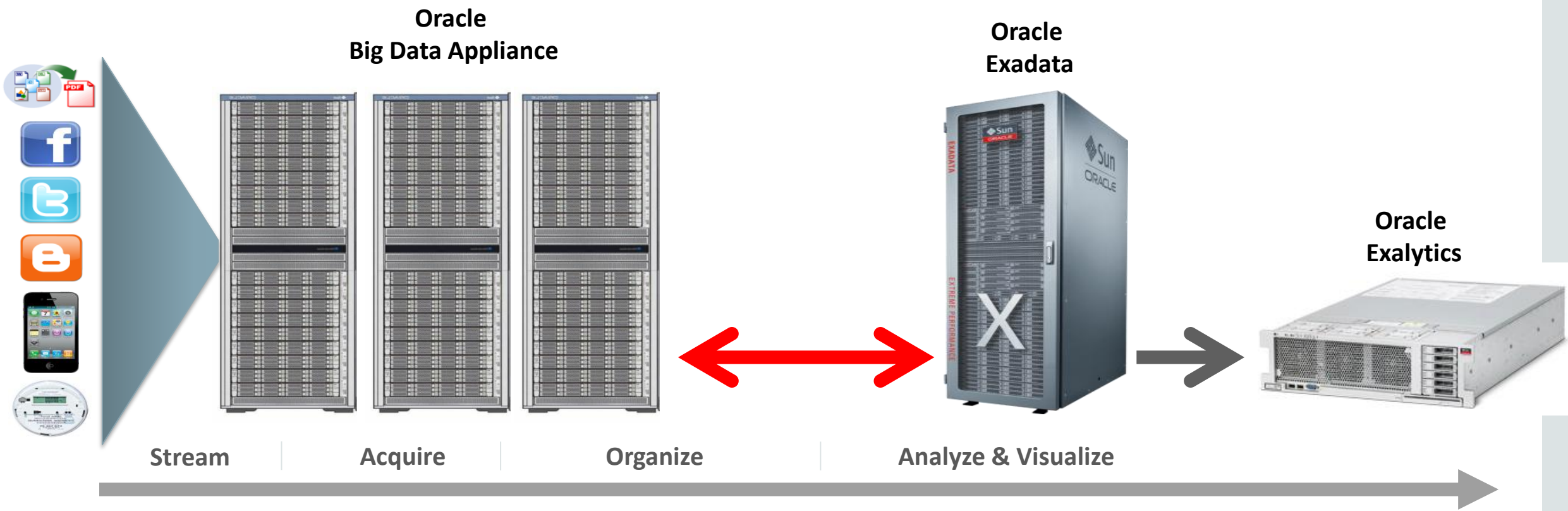


Exploding Data Volumes



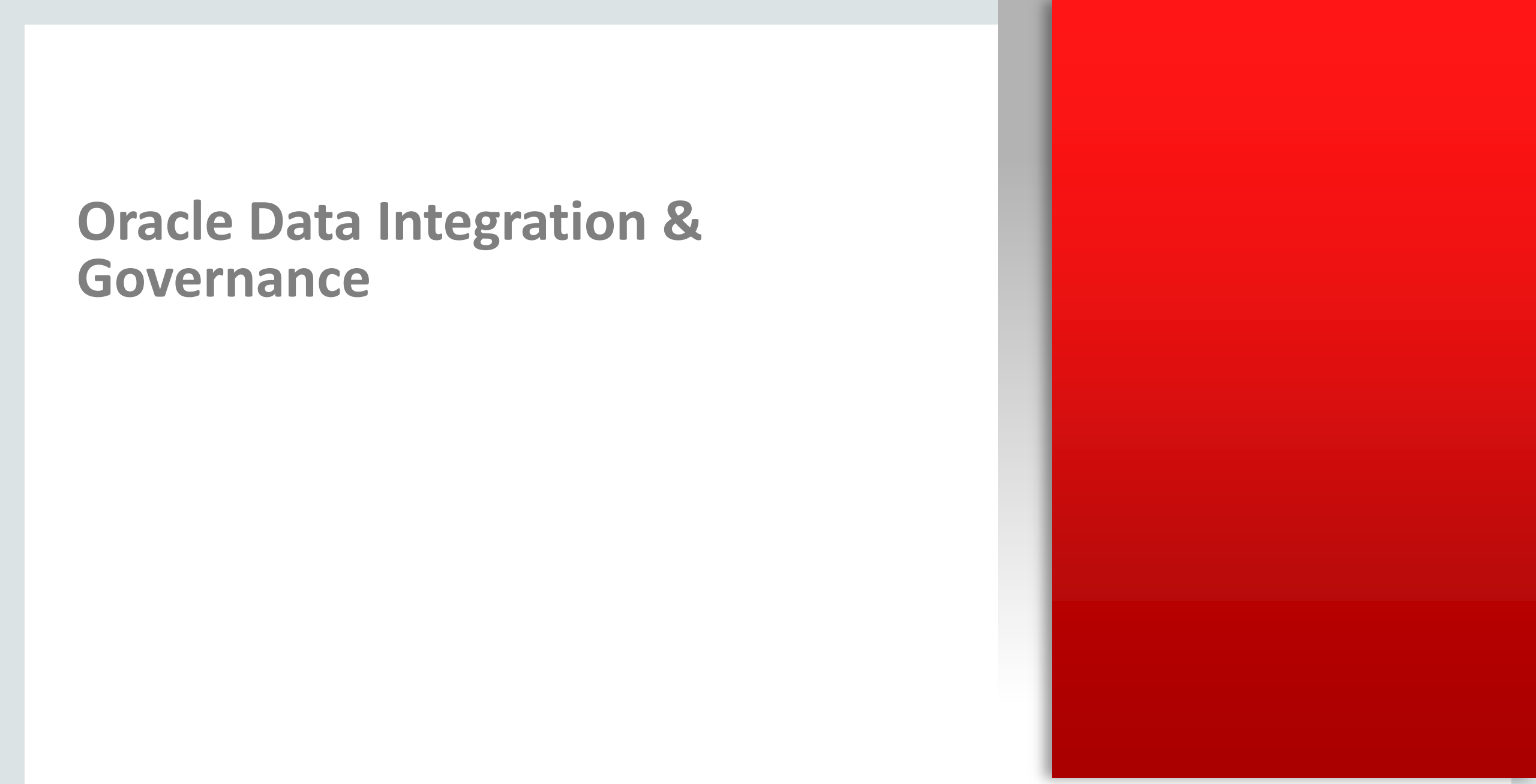
Source : Analysis Of Oracle Customer References

Big Data

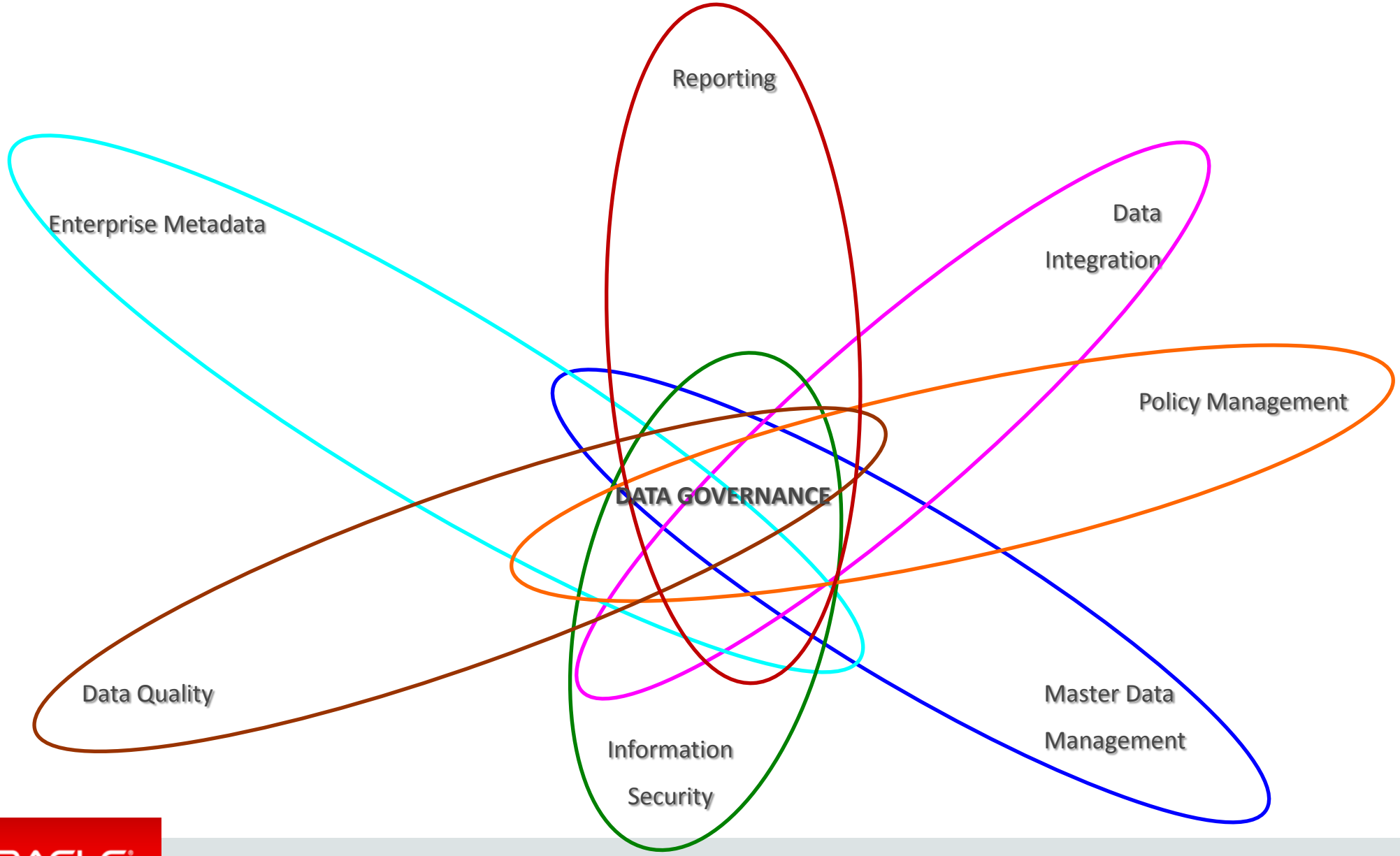


- *Load from big data processing into your data warehouse for further analysis*
- *Access your customer information while you process through your big data in order to look for patterns*

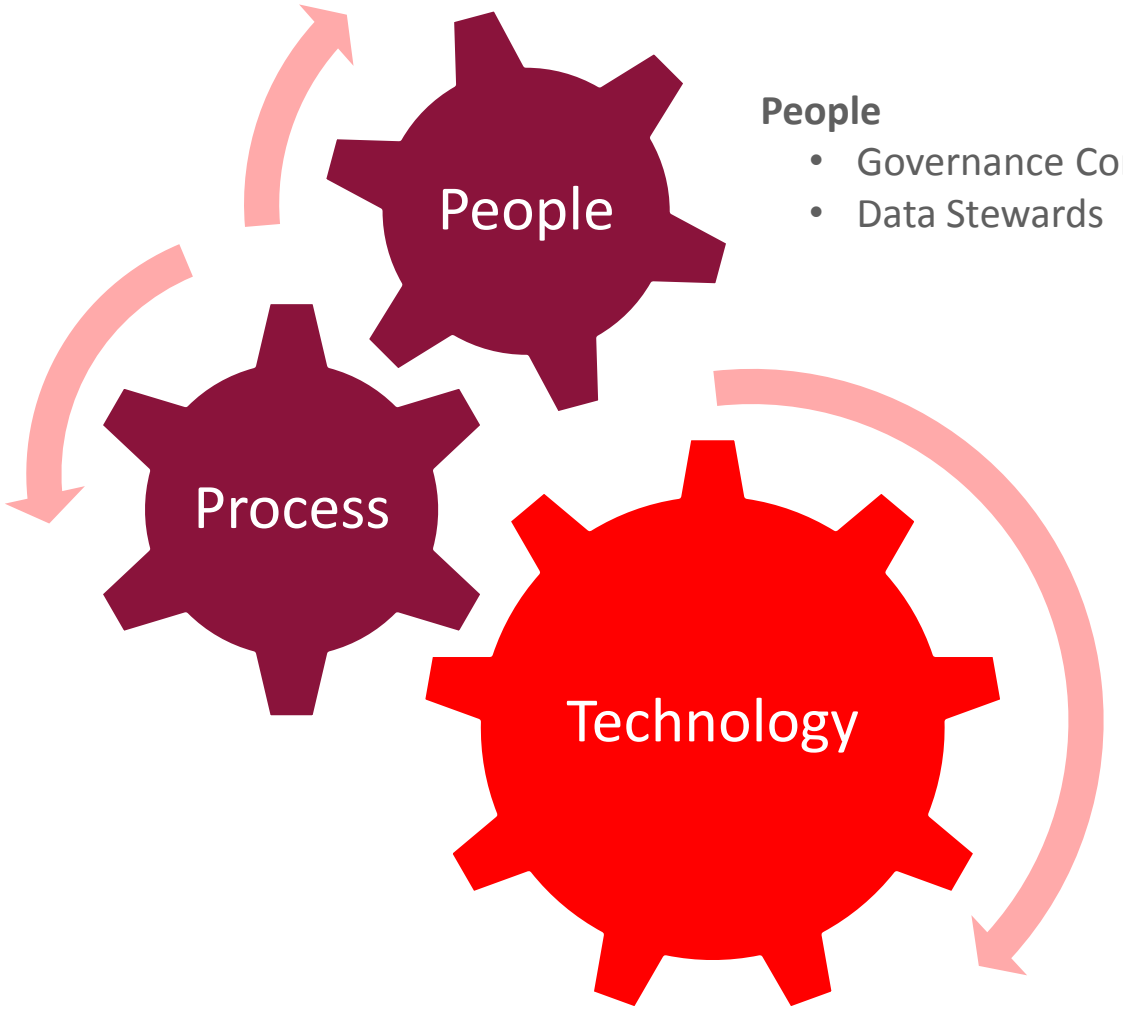
Oracle Data Integration & Governance



Data Governance Consolidates Strategies



Data Governance – more than just technology

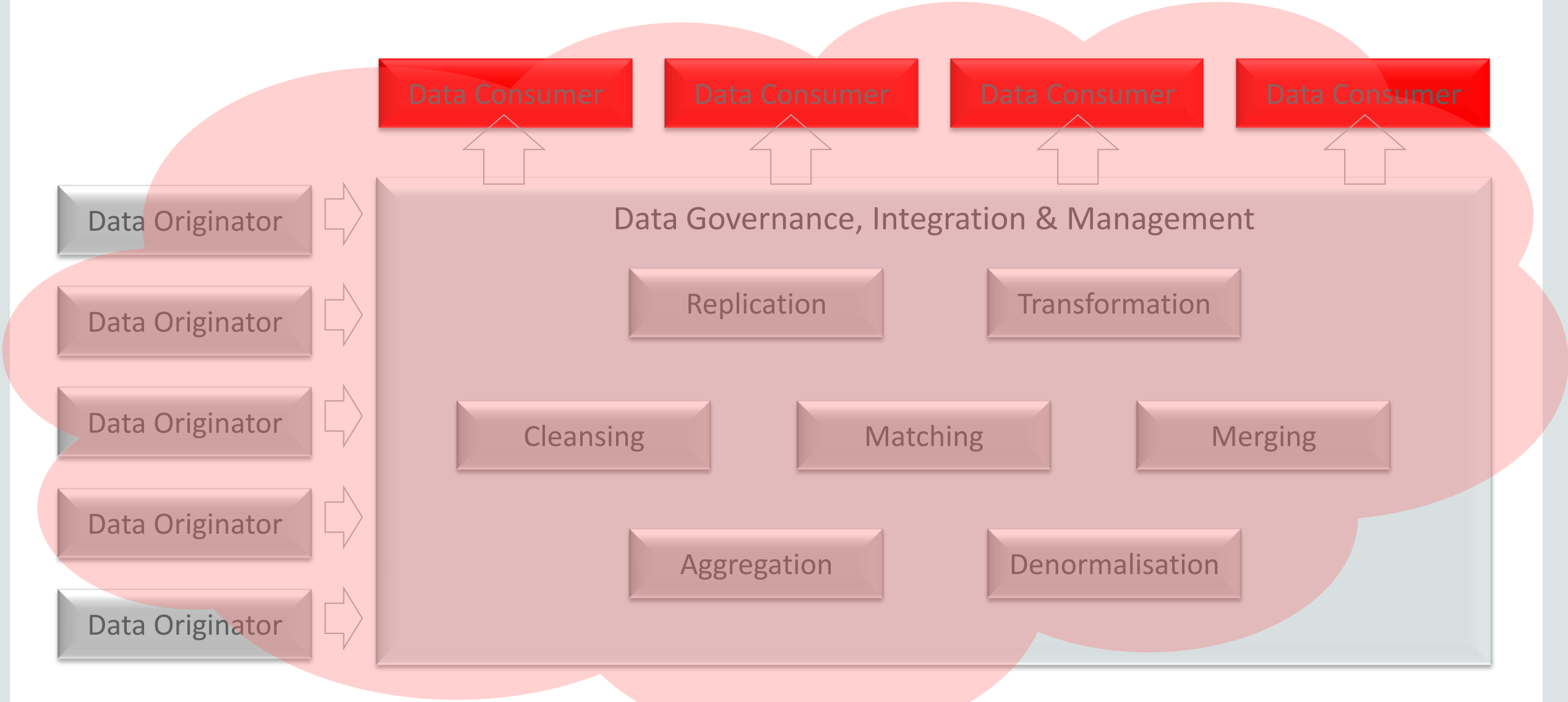


People
- Governance Committee
- Data Stewards

Process
- Defined rules & practices

Technology
- Implements process and supports people
- Makes operation more consistent & cost-effective
- Basis for continuous process improvement

Separation Between Originators and Consumers



Data Gov Economic Impact – Reduce Negatives

1-Reduce project risk by 40%

2-Avoid data remediation costs (manual effort, custom code) by 80% (Cut costs of handling duplicate data by 95%)

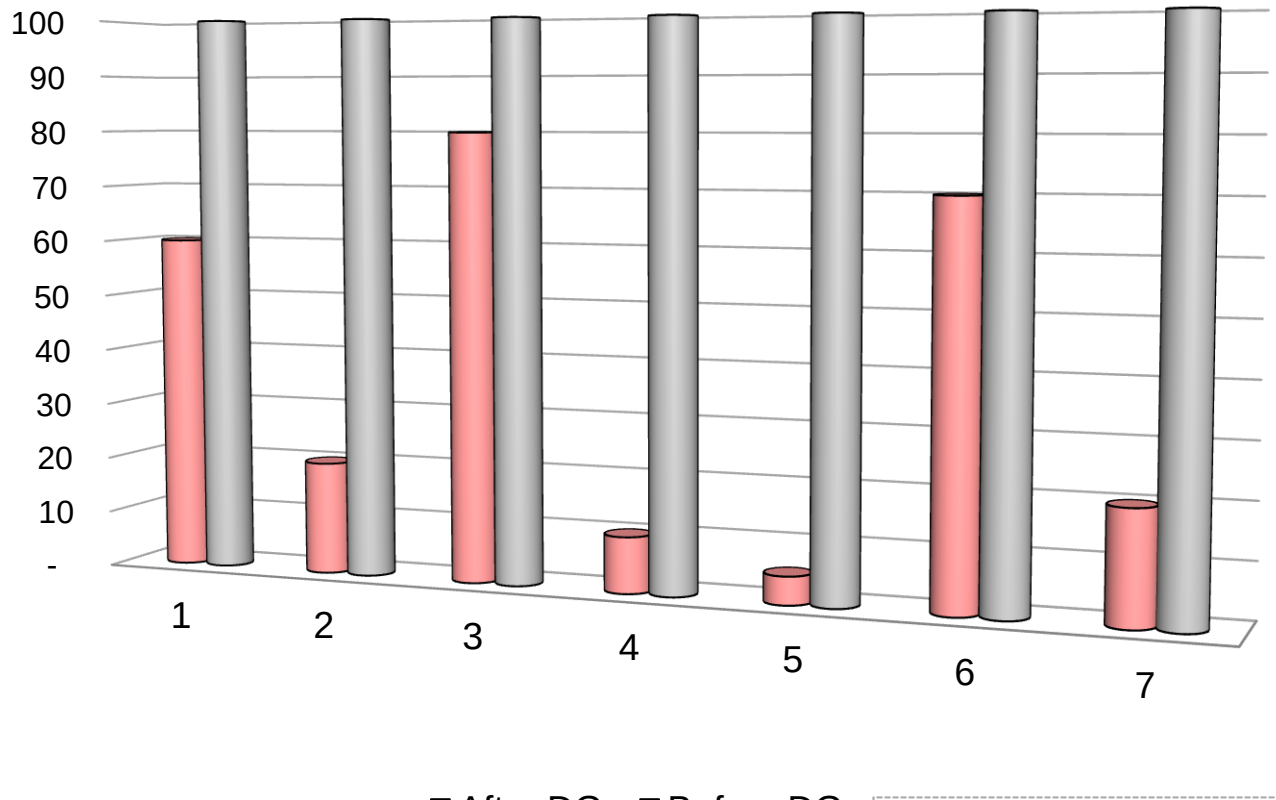
3-Avoid error costs (incorrect orders, inventory etc.) by 20%

4-Decrease costs of unnecessary system changes due to data quality problem by 90%

5- Avoid costs of un-doing work by 95%

6- Reduce analytical project costs by 30% (not anymore ..post remediation costs, unnecessary mistakes)

7 - Reduce data maintenance costs by 80% (not anymore.. hard to change, long to find)



Source : Analysis Of Oracle Customer References

Data Gov Economic Impact – Improve Positives

8-Speed up other systems, integration & processes by 10% (outputs and inputs in communication are correct and accurate data)

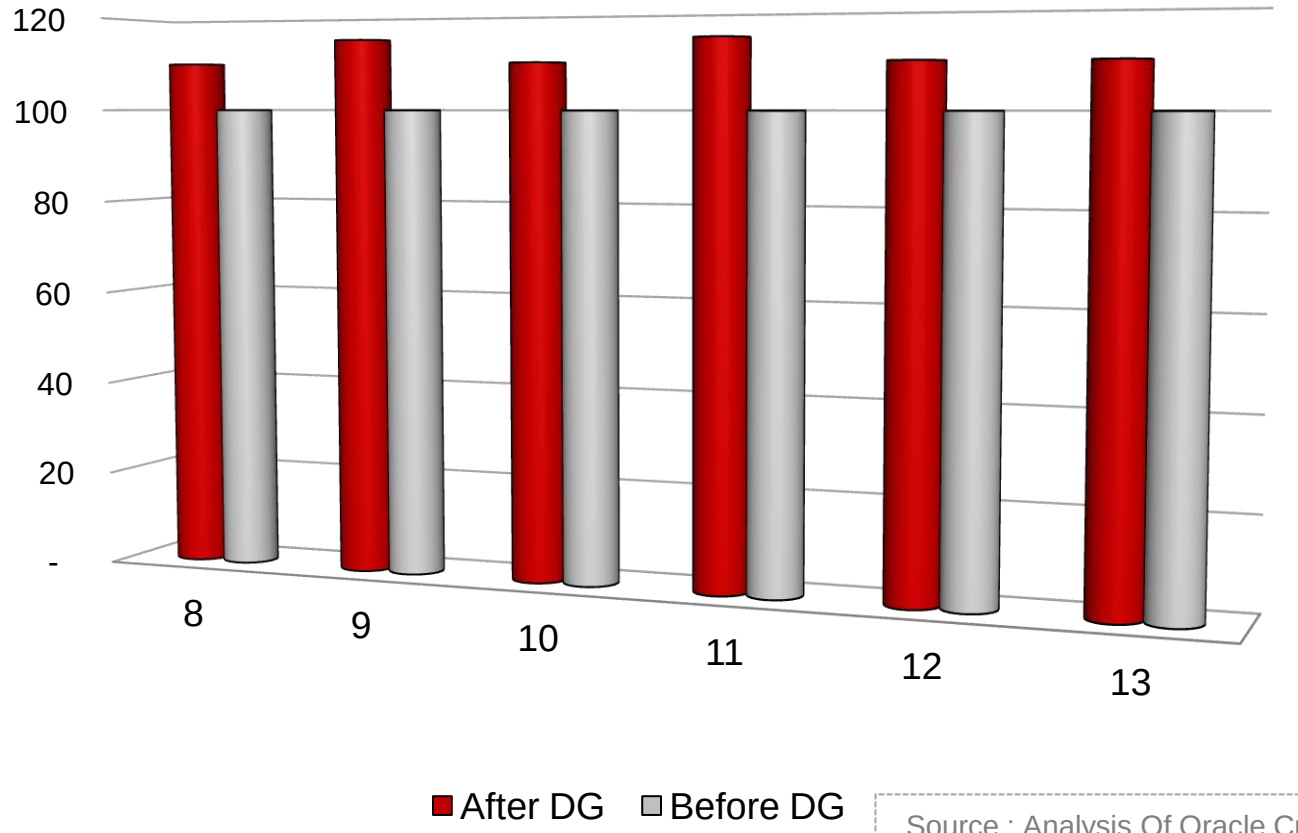
9-Increase BI, DW, CRM, Apps ROI by 15%

10-Gain by 10% more revenue due to agility to quickly reacting on ongoing market opportunities

11- Increase internal productivity and efficiency due to correct data by 15%

12- Increase revenue by 10% from gaining new customers and customer satisfaction Increase scalability

13- Improve the value of the company to those who would acquire it by 10%



Source : Analysis Of Oracle Customer References

DG Economic Impact

1-Reduce project risk by 40%

2-Avoid data remediation costs (manual effort, custom code) by 80% (Cut costs of handling duplicate data by 95%)

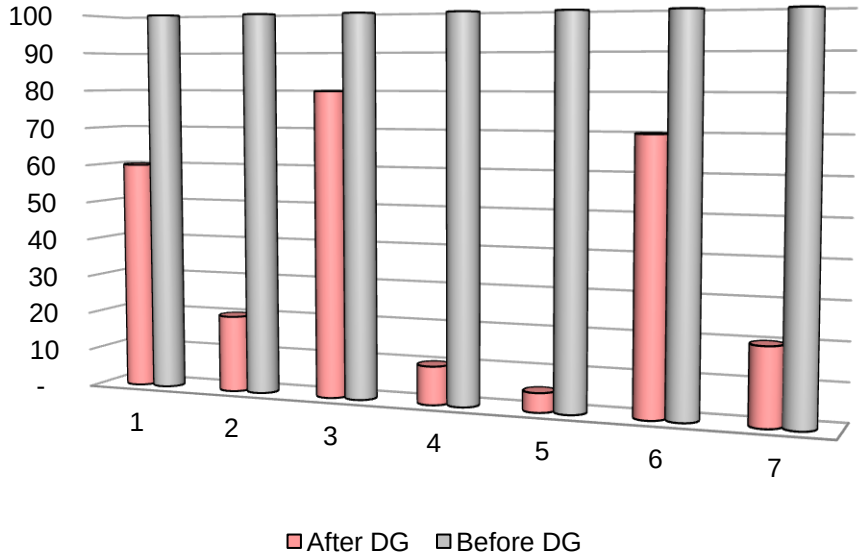
3-Avoid error costs (incorrect orders, inventory etc.) by 20%

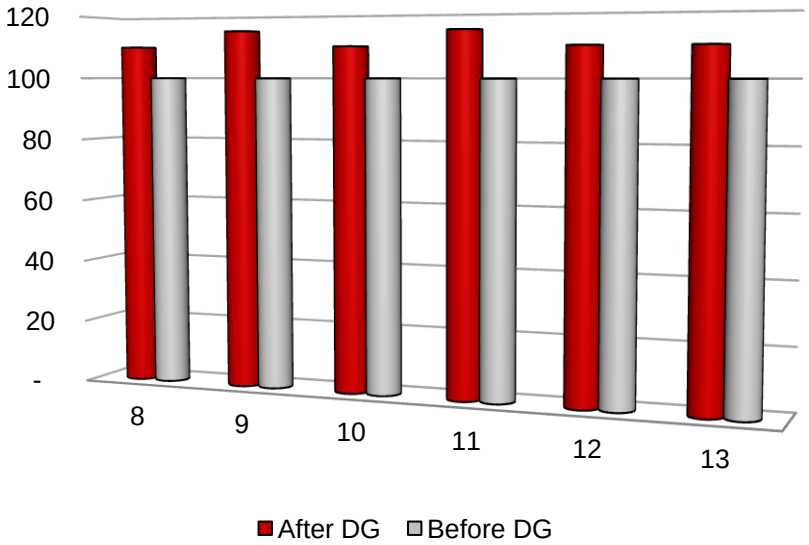
4-Decrease costs of unnecessary system changes due to data quality problem by 90%

5- Avoid costs of un-doing work by 95%

6- Reduce analytical project costs by 30% (not anymore ..post remediation costs, unnecessary mistakes)

7 - Reduce data maintenance costs by 80% (not anymore.. hard to change, long to find)





8-Speed up other systems, integration & processes by 10% (outputs and inputs in communication are correct and accurate data)

9-Increase BI, DW, CRM, Apps ROI by 15%

10-Gain by 10% more revenue due to agility to quickly reacting on ongoing market opportunities

11- Increase internal productivity and efficiency due to correct data by 15%

12- Increase revenue by 10% from gaining new customers and customer satisfaction

13- Improve the value of the company to those who would acquire it by 10%

ROI – 5 Areas To Demonstrate Value

(1) Program
Strategy & Direction
Structure
Scope
Prioritization
Policy
Roles & Responsibilities
Authority & Accountability
Issue Resolution
Business Alignment
Power To Question
Transparency
Weight & Balances
CoE
Agility To Respond

(2) Data Management Operations
Data Quality
Data Modeling
Metadata Management
Data Integration
Data Lineage
Security & Privacy
Data Standards
Business Rules
Business Process
Stewardship
Monitoring&Metrics
Reporting
Workflow Management

(3) Projects
MDM
BI
CRM
CDI
DW
EDW
DM
ERP
360 View
Data Migration
Data Integration

(4) Business Operations
Sales & Marketing
Customer Service
Risk & Compliance
Order To Cash
Supply Chain
Taxes & Fees
Social & Human Services
Mail & Transport
Asset Management
Business Analytics & Research

(5) Organization Strategy & Policy
Customer Definition
Merger & Acquisition
Planning & Budget
Business Partner Integration
Data Enrichment
Security & Privacy

ORACLE

Oracle Data Governance ROI

Increase revenue / value of assets

❑Improve the value of the company to those who would acquire it
❑Utilize information assets to make new sales, Better understand customers, Better understand product (and other) hierarchies

Reduce costs

❑Reduce duplicate data management processes (example: costs of data modeling, data administration, data quality)
❑Reduce errors and associated costs (in software development, report development, information interpretation) due to lack of understanding of data or poor quality data

Support Compliance While Reducing Costs

❑Avoid audit fees due to lack of confidence in “authoritative data”
❑Reduce costs of pre-audit testing

Support Impact Analysis

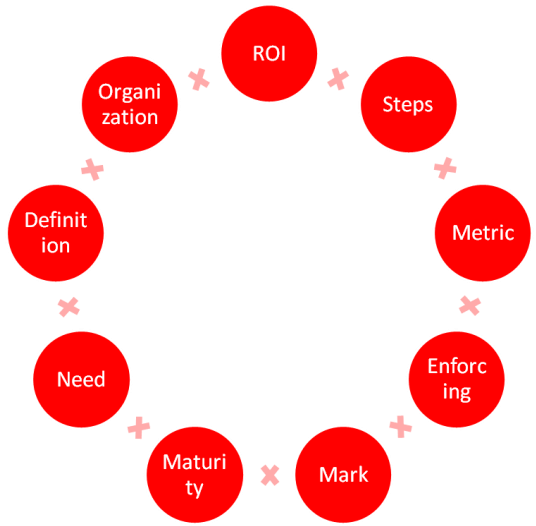
❑Provide a capability to assess cross-functional impacts of data-related decisions, to do useful impact analysis
(by providing authoritative business rules, system of record information, and data lineage metadata)
❑Avoid “undoing” work or rendering controls invalid

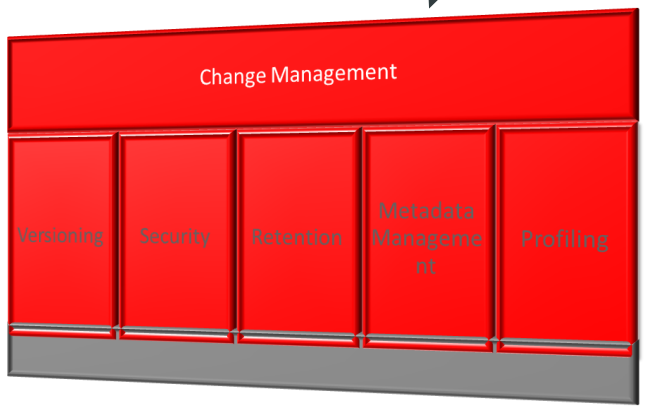
1

Òracle Data Governance -- User Experience

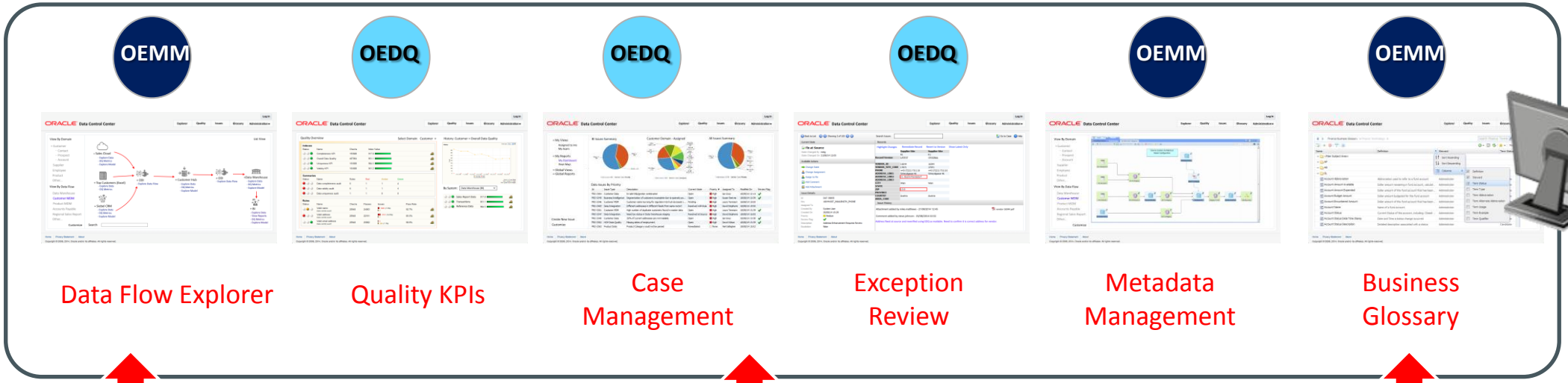
Data Flow Explorer | Quality KPIs | Case Management | Exception Review | Metadata Management | Business Glossary

2

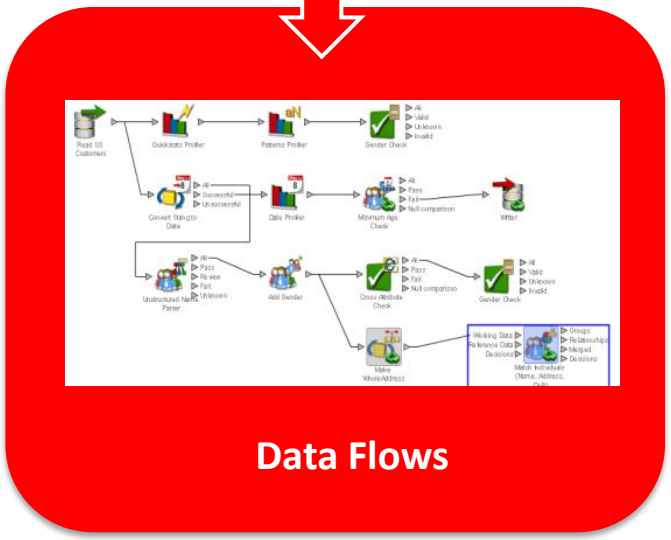




User Experience In Data Governance



Sources

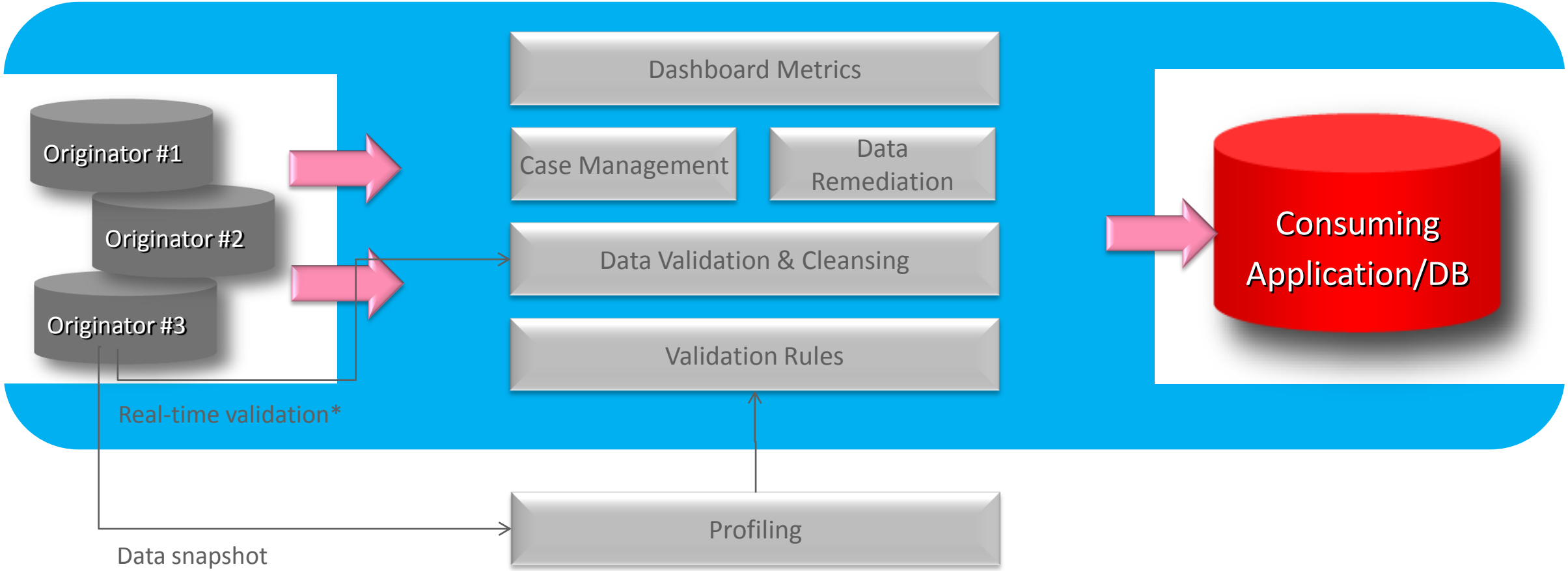


Targets

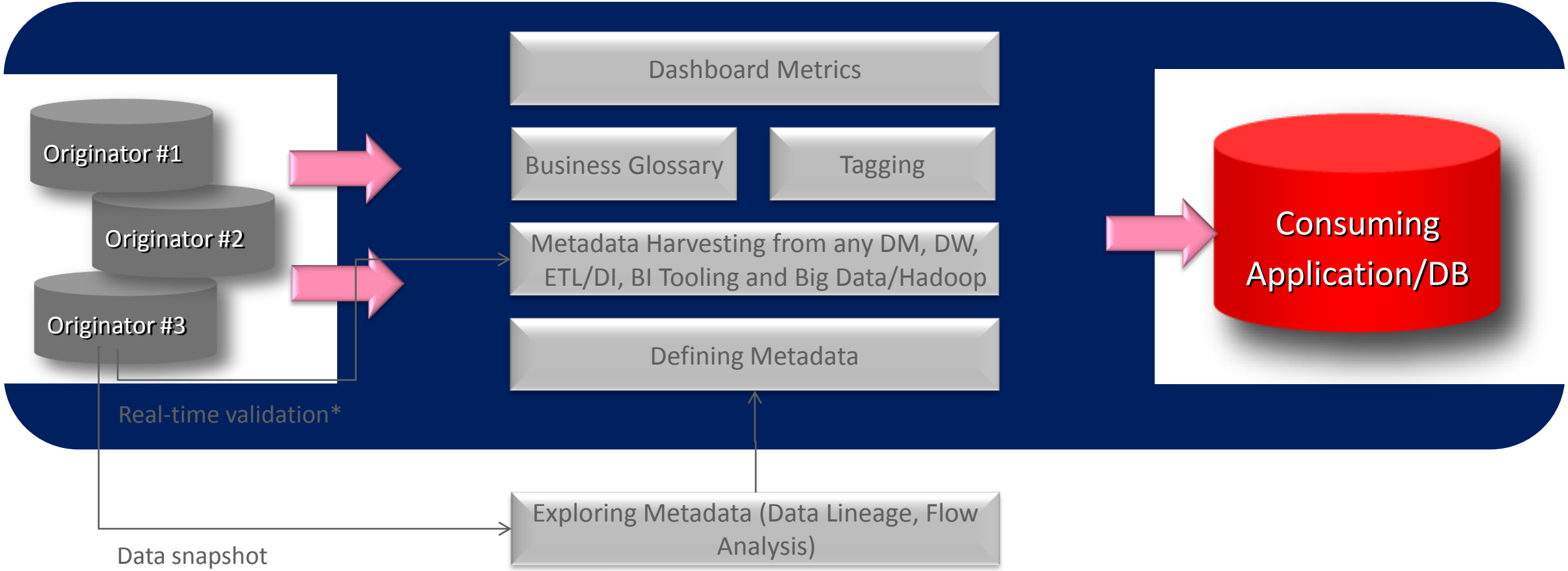
Oracle Enterprise Data Quality

Oracle Enterprise Metadata Management

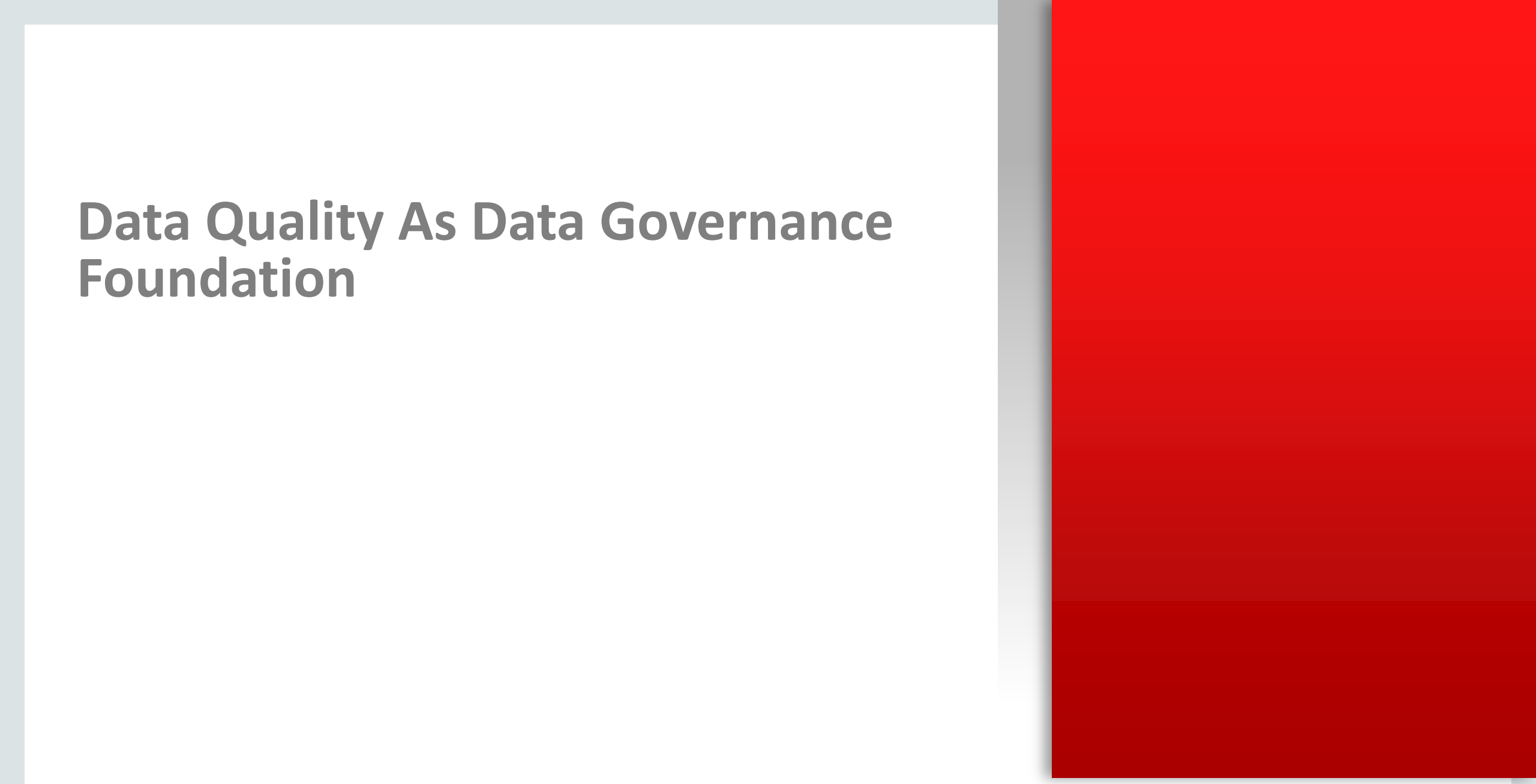
A First Step Of User Experience



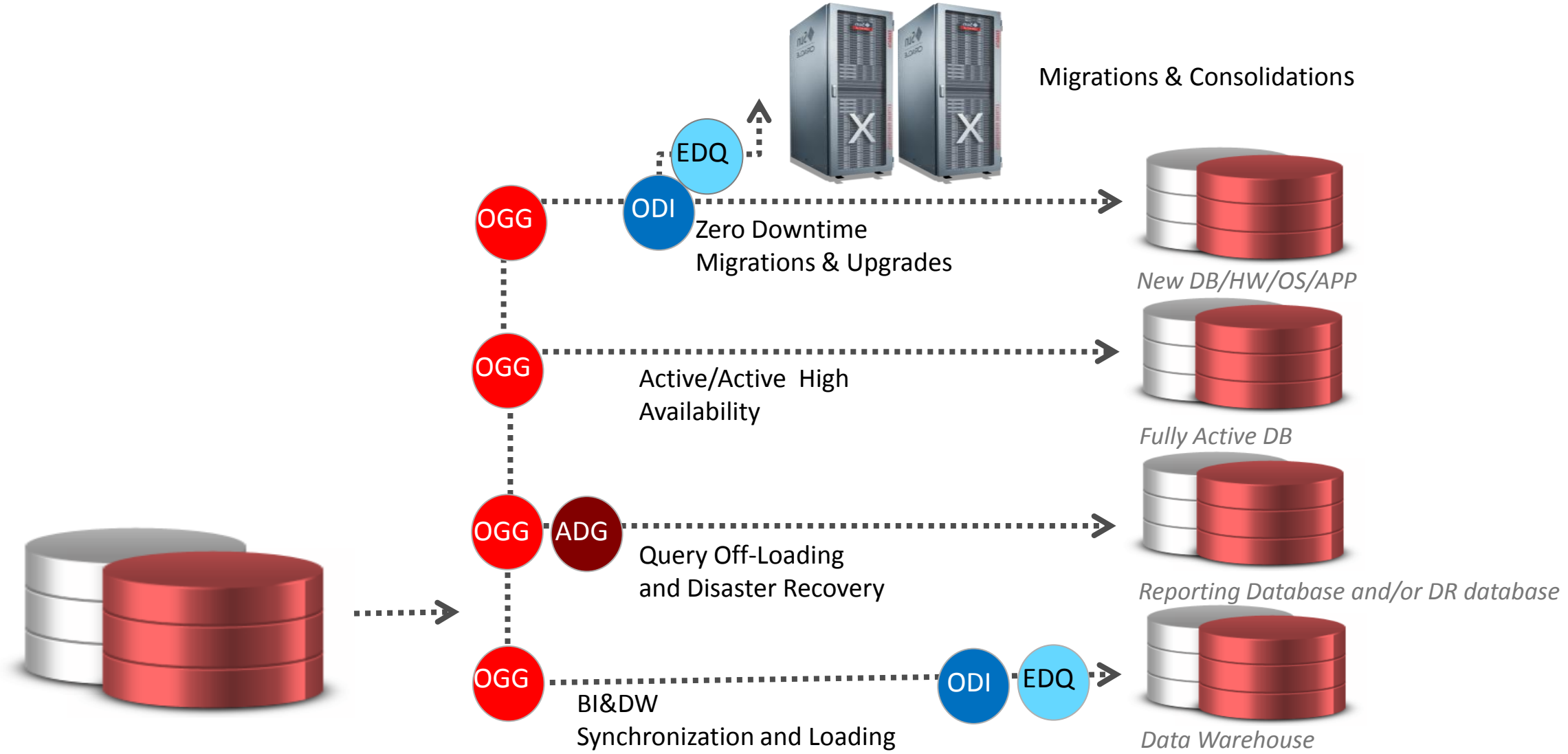
A Second Step Of User Experience



Data Quality As Data Governance Foundation

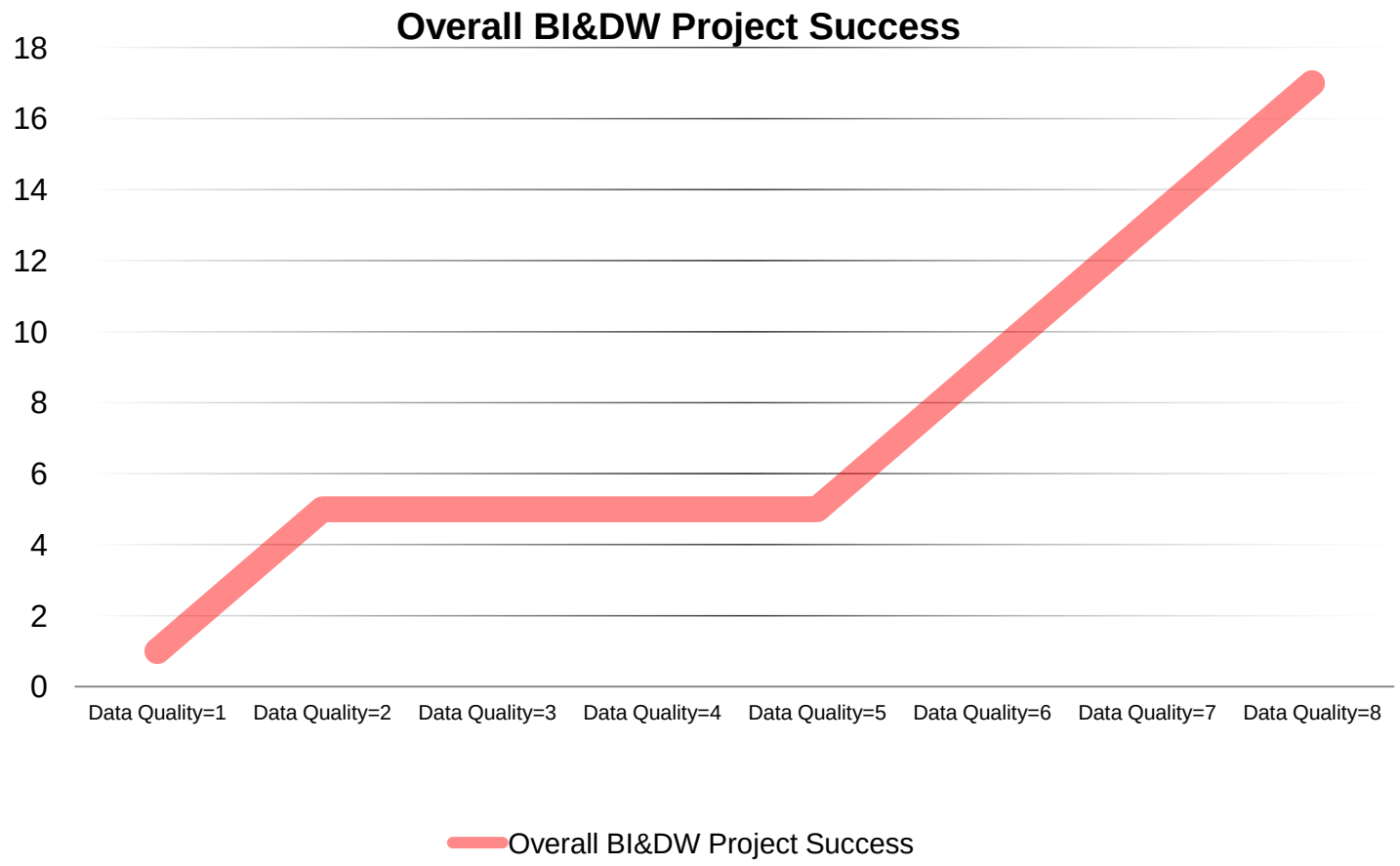


Use Cases Of EDQ



Failed BI&DW Projects

BI success is directly related to the quality of the underlying data from HC investment.



Why EDQ?

Inconsistent formats

Abbreviations (often ambiguous)

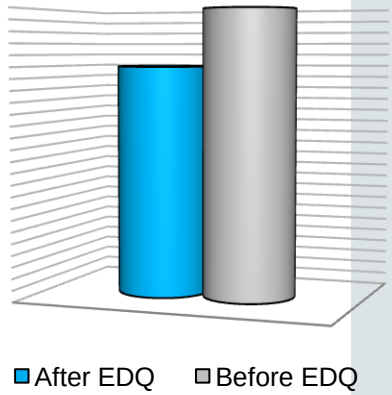
Attributes non-standard, missing or invalid

Customer ID	Customer Name	Address 1	Address 2	City	State	Zip	Country	Birth Date	Gender
AD23298	Mr Peter Mayhew	9407 Main St		Fairfax	VA	22031-4001	USA	02/23/61	M
VS38611	Dr Ellen Van Der Heijde (1)	144 E Grove St		Kingston	PA	18704	US	07/12/57	
DC18223	Jalila Abdul-Alim (Do Not Call) (2)	4548 Pennsylvania Ave	Apt 205	Kansas City	MO	64111-3349	USA	02/23/63	F
CO9387A	Tayside Computers Inc. (3)	4912 E 41st N		Idaho Falls	ID	83401	USA	31/03/2007 (7)	A
TZ35019	Mr Zachary P Jahn	98-1731 Ipuala Loop	Aiea	Hawaii	96701 (5)		United States	06/12/86	Male
CB27843	Mrs Edith Y Baba Junior	Baba Real Est. Corp. (3)	209 Stony Point Trl	Webster	NY		USA	11/17/1971	M (6)
OX80306	Andrew & Mary Baxter (4)	14 Oxbridge Way		Milfrod (6)	NH	03055-4614	US	05/28/67	F
JP70210	Mr RJ & Mrs FB MacDonald (4)	57 Hadleigh Close	Westlea	Swindon SN5 9BZ	MA	-	(6)	-	Y
RD48107	Mr Andy Baxter	14 Oxbridge Wy		Milford	NH	3056	USA	01/01/01 (8)	M

Widespread duplication (often hard to spot)

1. Compound Names
2. Embedded Additional Information
3. Mixed Business & Personal Names
4. Multiple Names
5. Mis-Fielded Data
6. Erroneous Data
7. International Date Formats
8. Default or Dummy Data

Avoid error costs (incorrect orders, inventory etc.) by 20%



Companies

In one hour…

- 240 businesses will change addresses
- 150 business telephone numbers will change or be disconnected
- 112 directorship (CEO, CFO, etc.) changes will occur
- 20 corporations will fail
- 12 new businesses will open their doors
- 4 companies will change their name

Individuals

In one hour…

- 5,769 individuals in the US will change jobs
- 2,748 individuals will change address
- 515 individuals will get married
- 263 individuals will get divorced
- 186 individuals will declare a personal bankruptcy

Products

In one year…

- On average 20% duplicates in product data
- 90% product introductions fail
- Retailers lost 40 billion or 3.5% of total sales lost each year due to item info inefficiencies
- 60% error rate for all invoices generated
- Global Data Sync will realize 30% lower IT costs

Master data changes at rate of 2% per month

Compounded, 2% monthly change is 27% per year, 61% in two years, *104%* in three years!!!

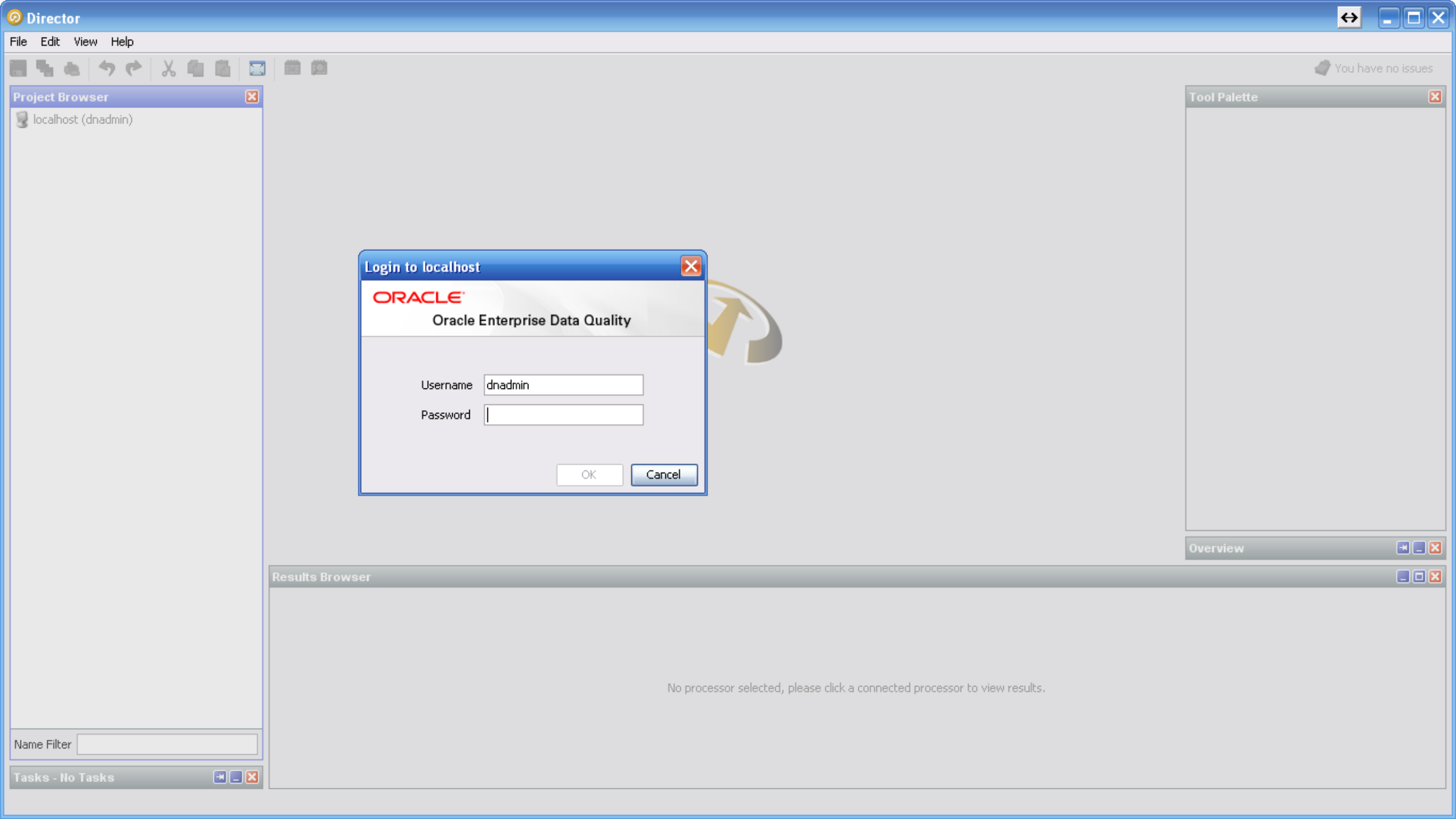
Source: D&B, US Census Bureau, US Department of Health and Human Services, Administrative Office of the US Courts, Bureau of Labor Statistics, Gartner, A.T Kearney, GMA Invoice Accuracy Study

ORACLE

Why EDQ?

Variation or Error	Example
Sequence errors	• Mark Douglas or Douglas Mark
Involuntary corrections	• Browne – Brown
Concatenated names	• Mary Anne, Maryanne
Nicknames and aliases	• Chris – Christine, Christopher, Tina
Noise	• Full stops, dashes, slashes, titles, apostrophes
Abbreviations	• Wlm/William, Mfg/Manufacturing
Truncations	• Credit Suisse First Bost
Prefix/suffix errors	• MacDonald/McDonald/Donald
Spelling & typing errors	• P0rter, Beht

Variation or Error	Example
Transcription mistakes	• Hannah, Hamah
Missing or extra tokens	• George W Smith, George Smith, Smith
Foreign sourced data	• Khader AL Ghamdi, Khadir A. AlGamdey
Unpredictable use of initials	• John Alan Smith, J A Smith
Transposed characters	• Johnson, Jhonson
Localization	• Stanislav Milosovich – Stan Milo
Inaccurate dates	• 12/10/1915, 21/10/1951, 10121951, 00001951
Transliteration differences	• Gang, Kang, Kwang
Phonetic errors	• Graeme – Graham


Director
File
Edit
View
Help
You have no issues
Project Browser
localhost (dnadmin)
Name Filter
Tasks - No Tasks
Tool Palette
Overview
Results Browser
No processor selected, please click a connected processor to view results.
Login to localhost
ORACLE
Oracle Enterprise Data Quality
Username
dnadmin
Password
OK
Cancel

Director

File Edit View Help

You have no issues

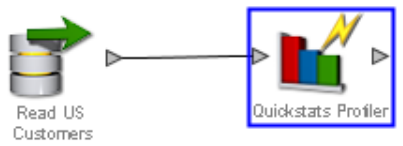
Project Browser

- localhost (dnadmin)
  - Projects
    - CDEP Processors on canvas
    - Compass Data Quality
    - Customer data demo - US
    - Demo
      - Data Stores
        - DemoData
      - Staged Data
        - US Customers
      - Views
      - Processes
        - Exploring US Customers
      - Reference Data
      - Results Books
      - External Tasks
      - Jobs
      - Exports
      - Web Services
      - Notes
    - Web Service demo
  - Reference Data
  - Data Stores
  - Published Processors

Name Filter

Tasks - No Tasks

Exploring US Customers



Process

Tool Palette - Profiling

- Date Profiler
- Equal Attributes Profiler
- Frequency Profiler
- Length Profiler
- Max/Min Profiler
- Number Profiler
- Patterns Profiler
- Quickstats Profiler
- Record Completeness Profiler
- Record Duplication Profiler
- RegEx Patterns Profiler

Search

Overview

Results Browser

Job: Exploring US Customers

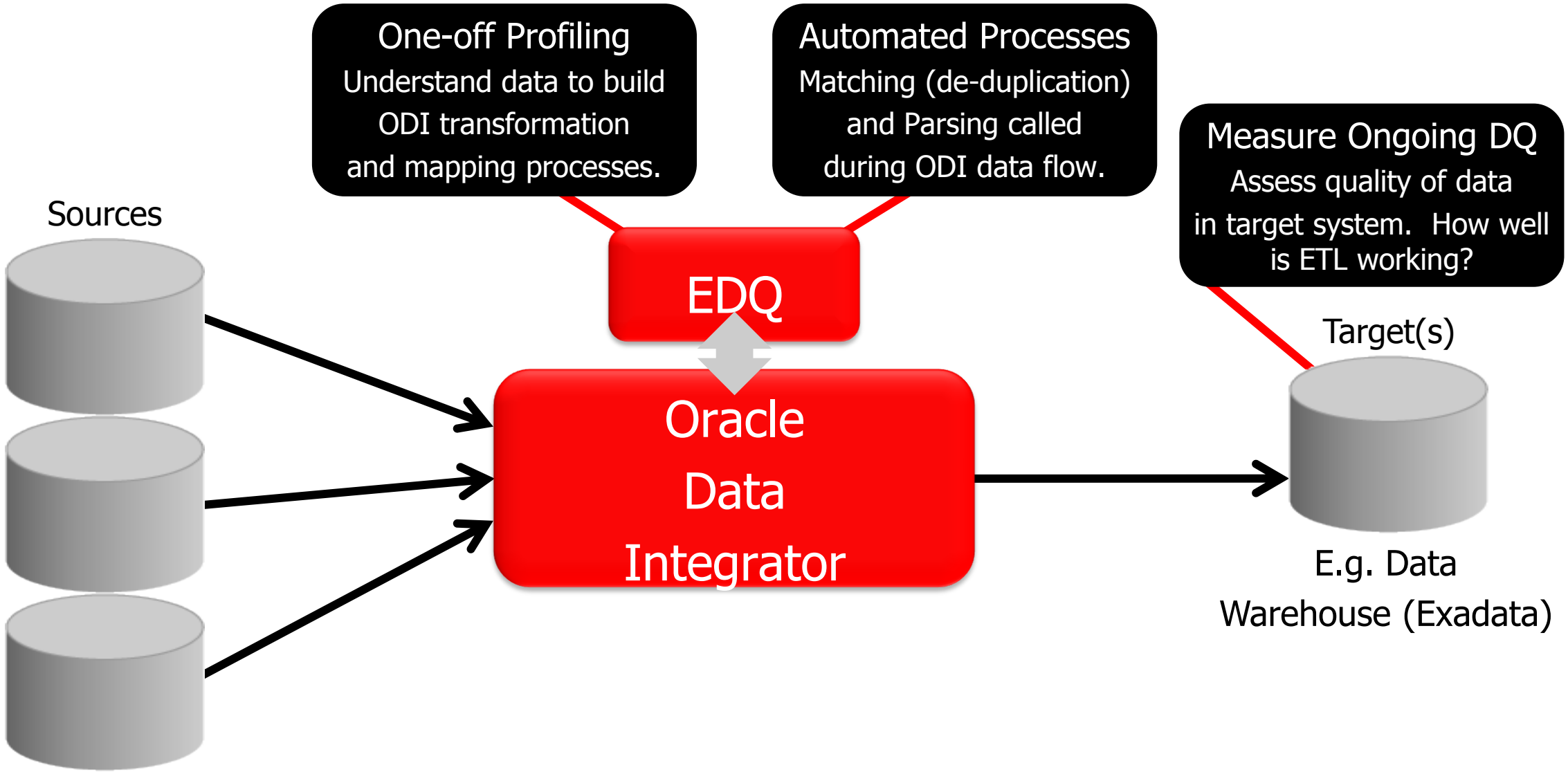
Latest Run: 20-Aug-2011 15:12:02 - 15:12:17

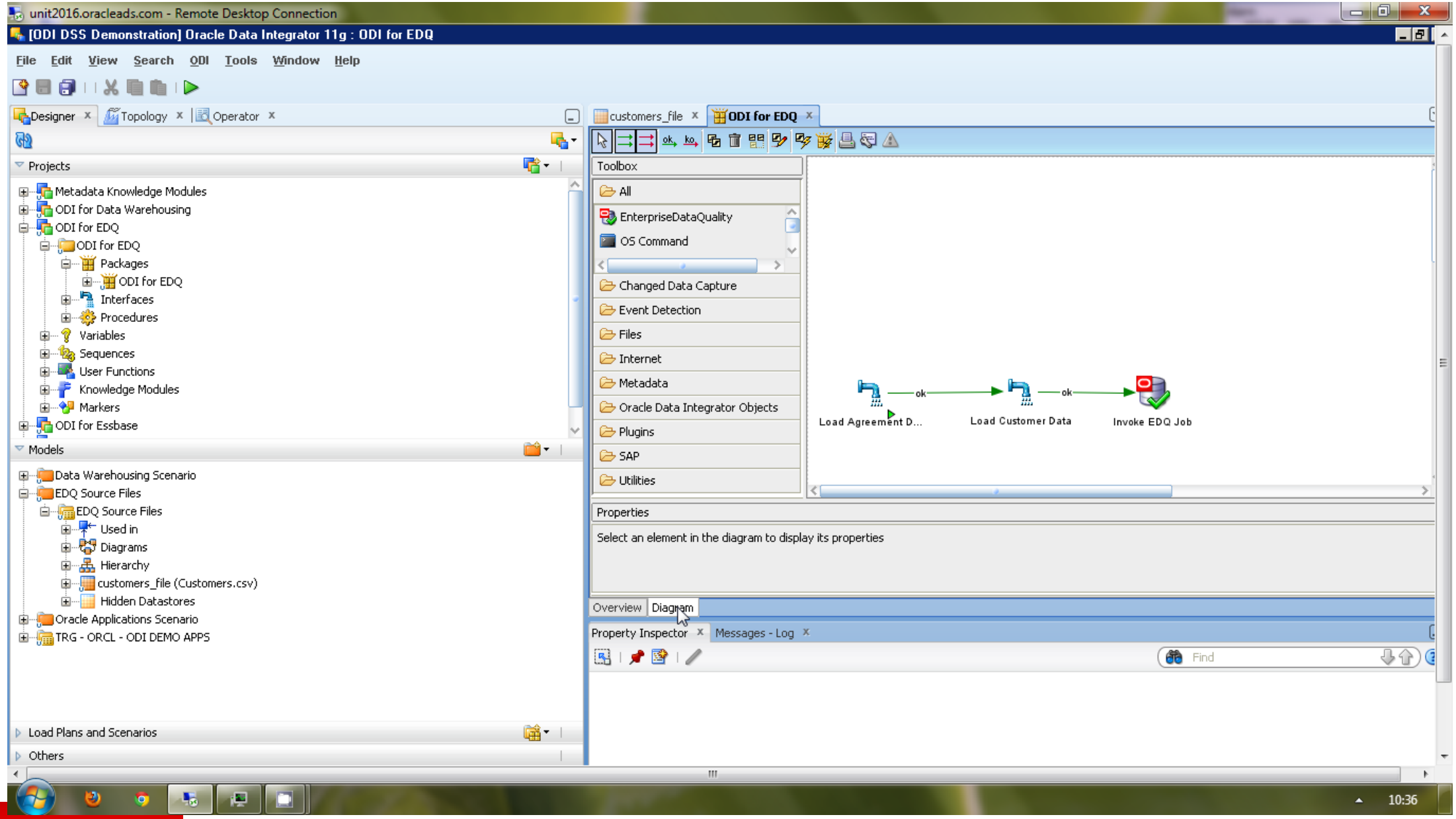
Viewing all 17 records

Input Field	Record Total	With Data	Without Data	Singleton	Duplicates	Distinct Values	Comment
ID	5438	5438	0	5438	0	5438	Complete; Possible key
Name	5438	5438	0	5327	111	5380	Complete; Potentially damaged key; Investigate duplicates
Street	5438	5438	0	5319	119	5376	Complete; Potentially damaged key; Investigate duplicates
City	5438	5438	0	396	5042	1232	Complete
State	5438	5438	0	12	5426	65	Complete
ZIP	5438	5436	2	490	4948	1823	Investigate blanks

Summary statistics view | Data

ODI&EDQ





Director

File Edit View Help

You have no issues

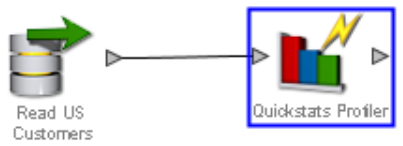
Project Browser

- localhost (dnadmin)
  - Projects
    - CDEP Processors on canvas
    - Compass Data Quality
    - Customer data demo - US
    - Demo
      - Data Stores
        - DemoData
      - Staged Data
        - US Customers
      - Views
      - Processes
        - Exploring US Customers
      - Reference Data
      - Results Books
      - External Tasks
      - Jobs
      - Exports
      - Web Services
      - Notes
    - Web Service demo
  - Reference Data
  - Data Stores
  - Published Processors

Name Filter

Tasks - No Tasks

Exploring US Customers



Process

Tool Palette - Profiling

- Date Profiler
- Equal Attributes Profiler
- Frequency Profiler
- Length Profiler
- Max/Min Profiler
- Number Profiler
- Patterns Profiler
- Quickstats Profiler
- Record Completeness Profiler
- Record Duplication Profiler
- RegEx Patterns Profiler

Search

Overview

Results Browser

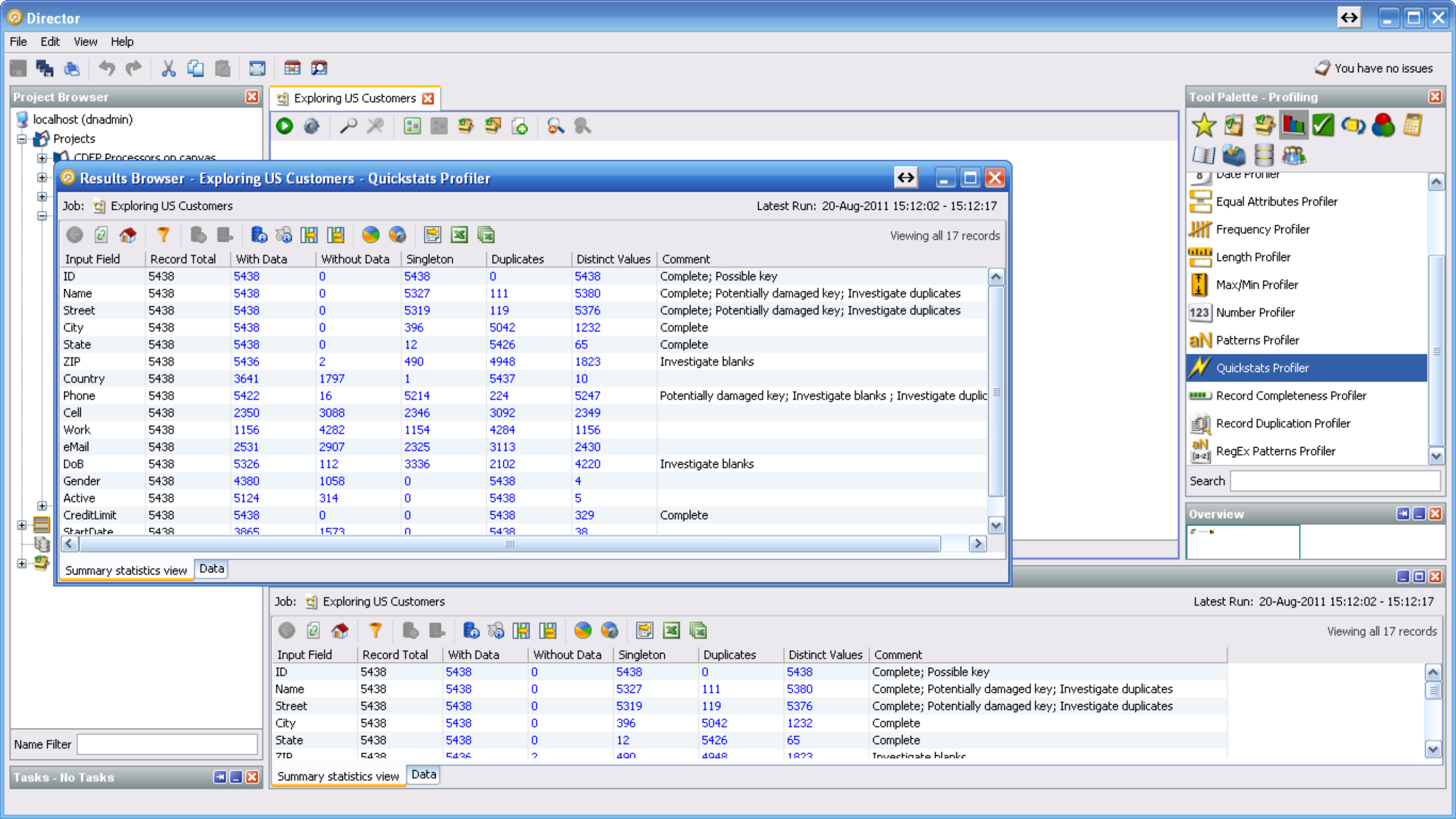
Job: Exploring US Customers

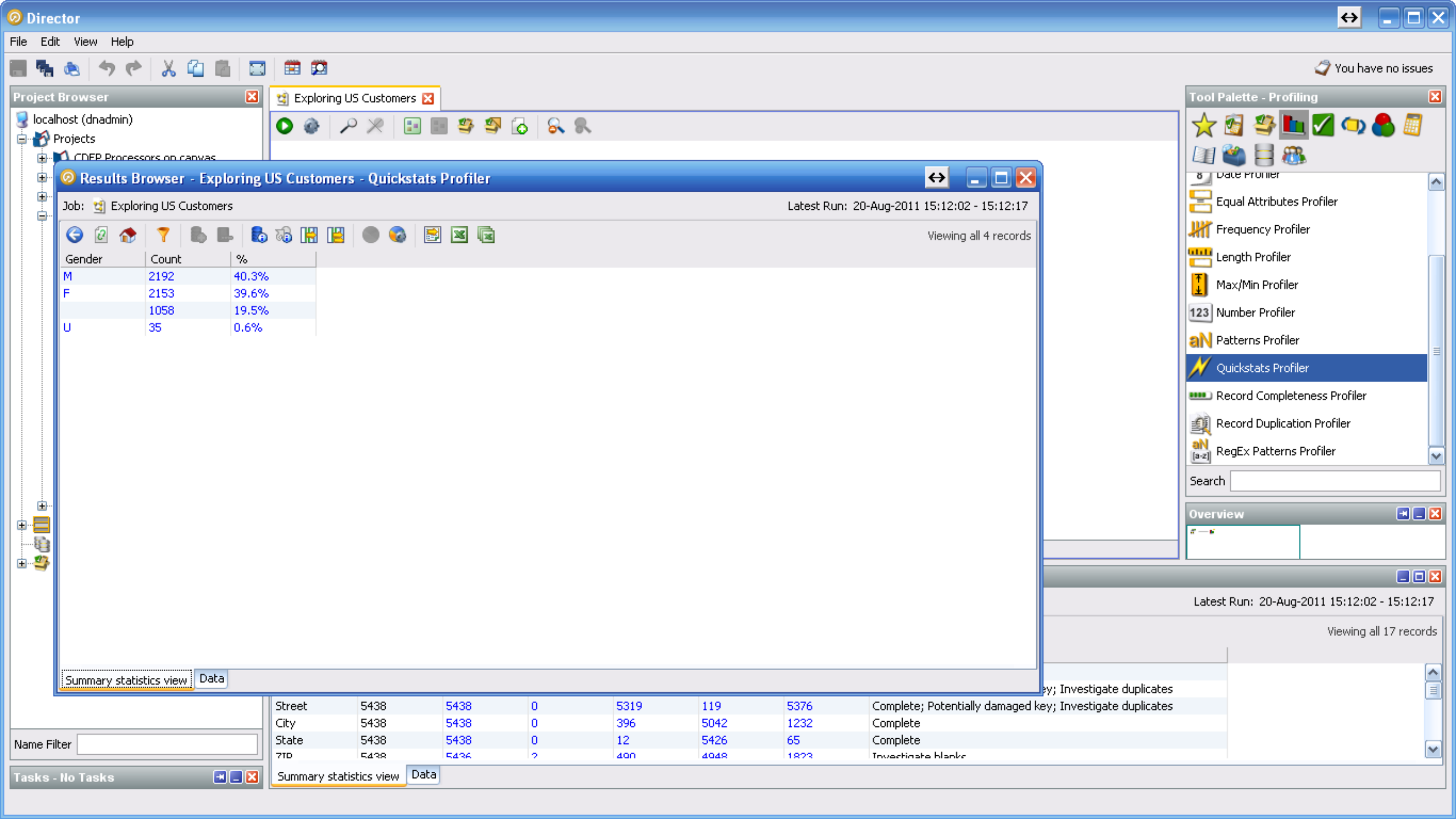
Latest Run: 20-Aug-2011 15:12:02 - 15:12:17

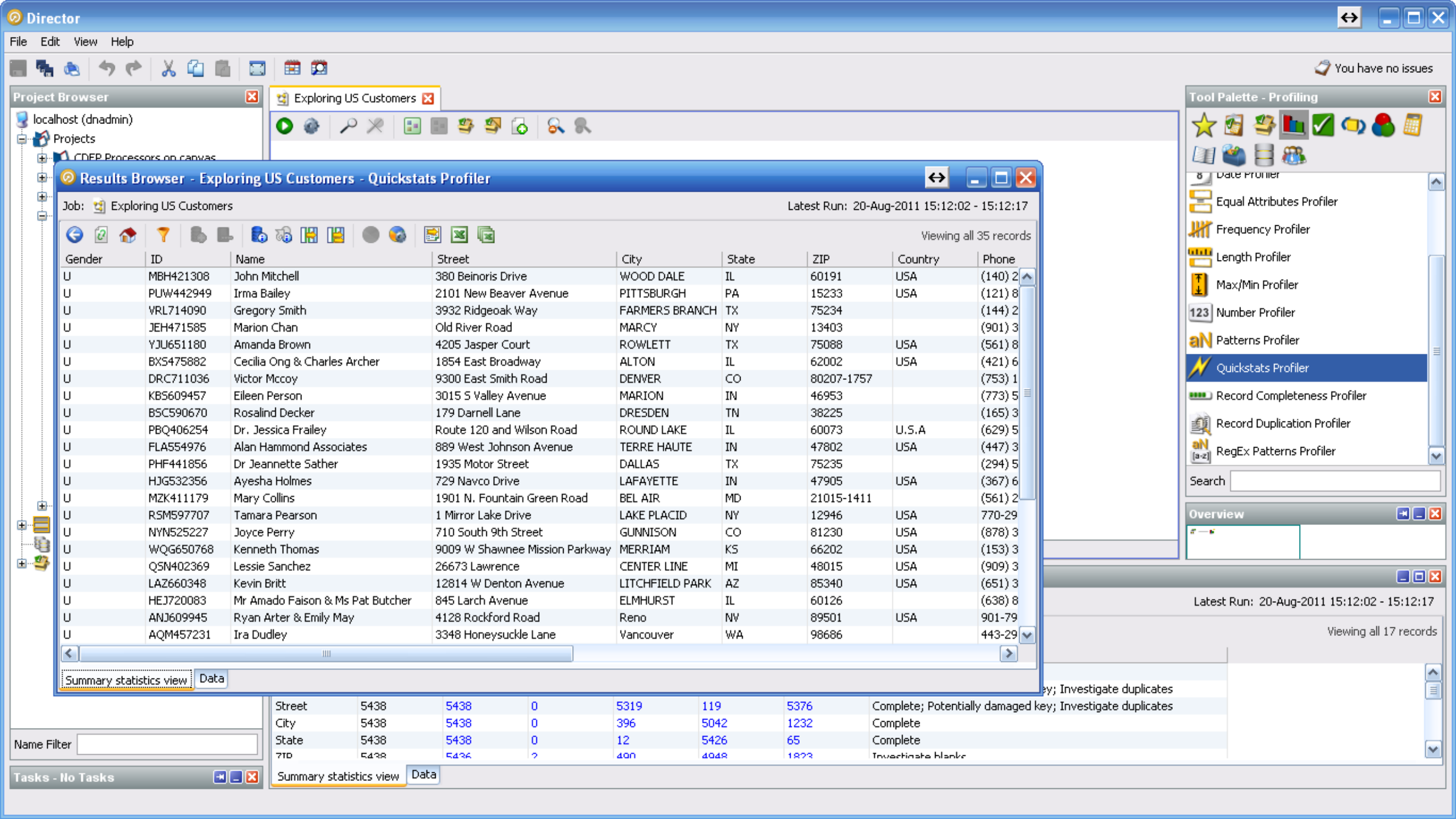
Viewing all 17 records

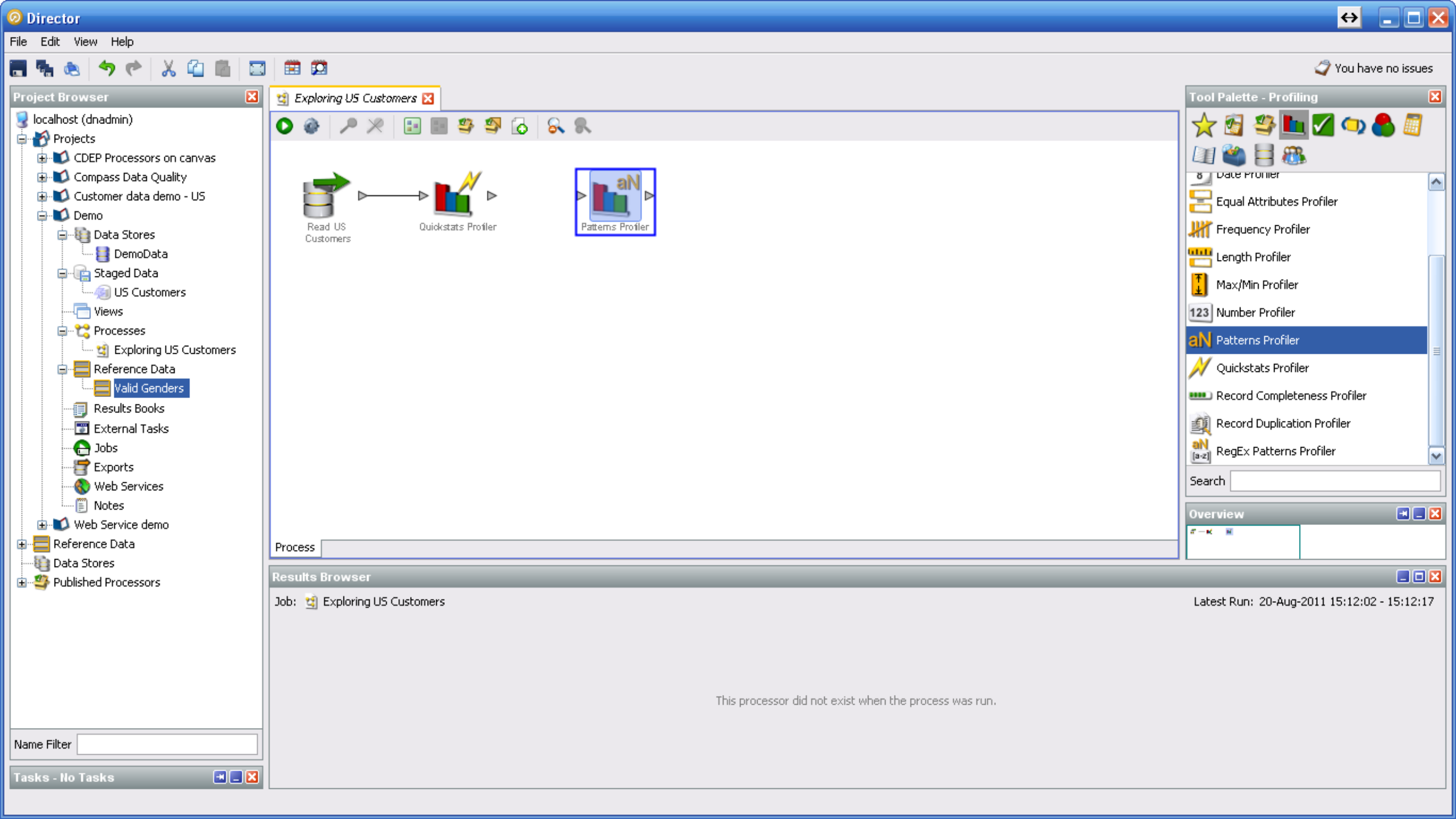
Input Field	Record Total	With Data	Without Data	Singleton	Duplicates	Distinct Values	Comment
ID	5438	5438	0	5438	0	5438	Complete; Possible key
Name	5438	5438	0	5327	111	5380	Complete; Potentially damaged key; Investigate duplicates
Street	5438	5438	0	5319	119	5376	Complete; Potentially damaged key; Investigate duplicates
City	5438	5438	0	396	5042	1232	Complete
State	5438	5438	0	12	5426	65	Complete
ZIP	5438	5436	2	490	4948	1823	Investigate blanks

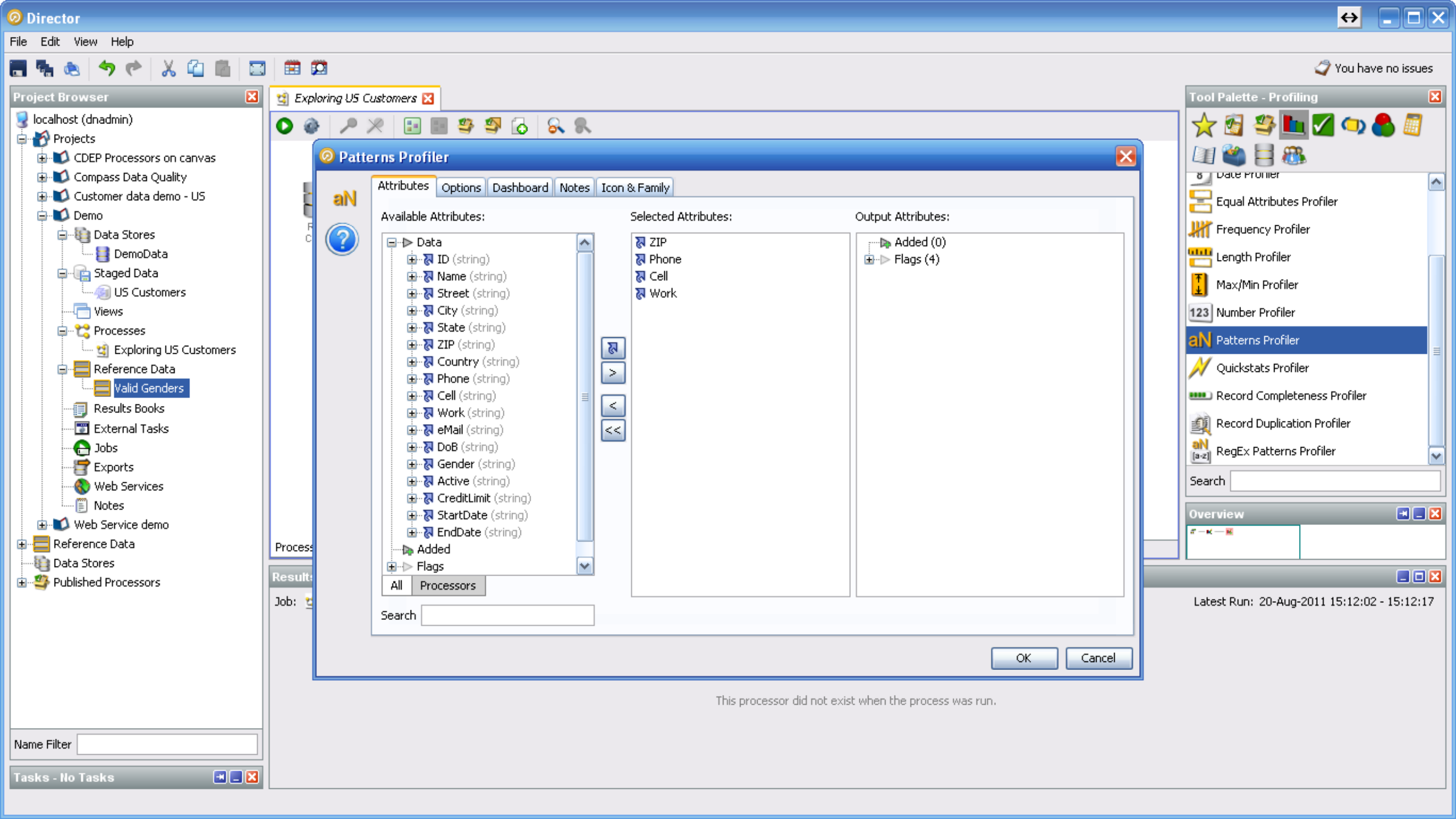
Summary statistics view | Data

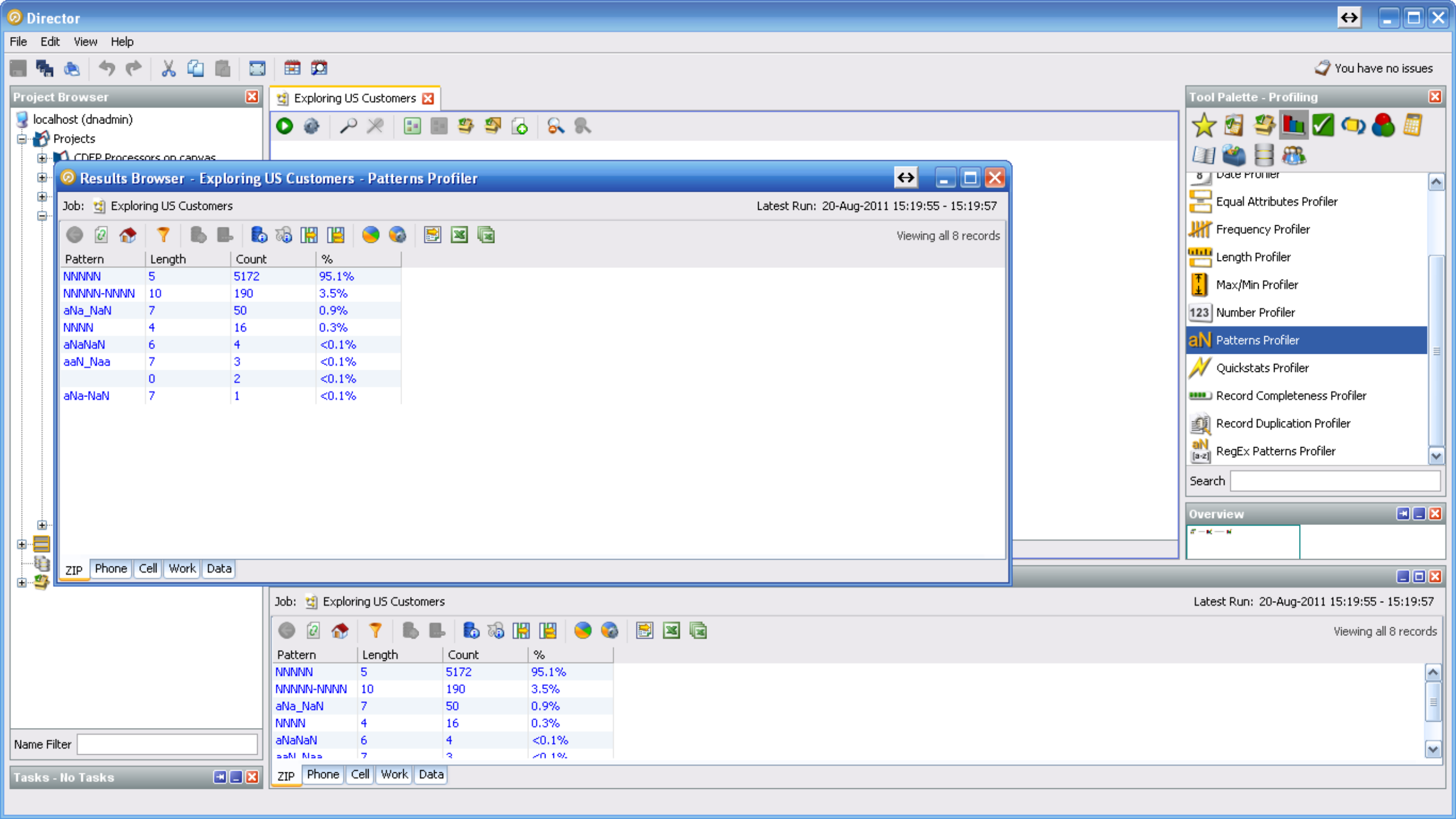

Director
File Edit View Help
You have no issues
Project Browser
localhost (dnadmin)
Projects
Exploring US Customers
Tool Palette - Profiling
Date Profiler
Equal Attributes Profiler
Frequency Profiler
Length Profiler
Max/Min Profiler
Number Profiler
Patterns Profiler
Quickstats Profiler
Record Completeness Profiler
Record Duplication Profiler
RegEx Patterns Profiler
Search
Overview
Results Browser - Exploring US Customers - Quickstats Profiler
Job: Exploring US Customers
Latest Run: 20-Aug-2011 15:12:02 - 15:12:17
Viewing all 17 records
Input Field | Record Total | With Data | Without Data | Singleton | Duplicates | Distinct Values | Comment
ID | 5438 | 5438 | 0 | 5438 | 0 | 5438 | Complete; Possible key
Name | 5438 | 5438 | 0 | 5327 | 111 | 5380 | Complete; Potentially damaged key; Investigate duplicates
Street | 5438 | 5438 | 0 | 5319 | 119 | 5376 | Complete; Potentially damaged key; Investigate duplicates
City | 5438 | 5438 | 0 | 396 | 5042 | 1232 | Complete
State | 5438 | 5438 | 0 | 12 | 5426 | 65 | Complete
ZIP | 5438 | 5436 | 2 | 490 | 4948 | 1823 | Investigate blanks
Country | 5438 | 3641 | 1797 | 1 | 5437 | 10
Phone | 5438 | 5422 | 16 | 5214 | 224 | 5247 | Potentially damaged key; Investigate blanks ; Investigate duplic
Cell | 5438 | 2350 | 3088 | 2346 | 3092 | 2349
Work | 5438 | 1156 | 4282 | 1154 | 4284 | 1156
eMail | 5438 | 2531 | 2907 | 2325 | 3113 | 2430
DoB | 5438 | 5326 | 112 | 3336 | 2102 | 4220 | Investigate blanks
Gender | 5438 | 4380 | 1058 | 0 | 5438 | 4
Active | 5438 | 5124 | 314 | 0 | 5438 | 5
CreditLimit | 5438 | 5438 | 0 | 0 | 5438 | 329 | Complete
Summary statistics view
Data
Name Filter
Tasks - No Tasks

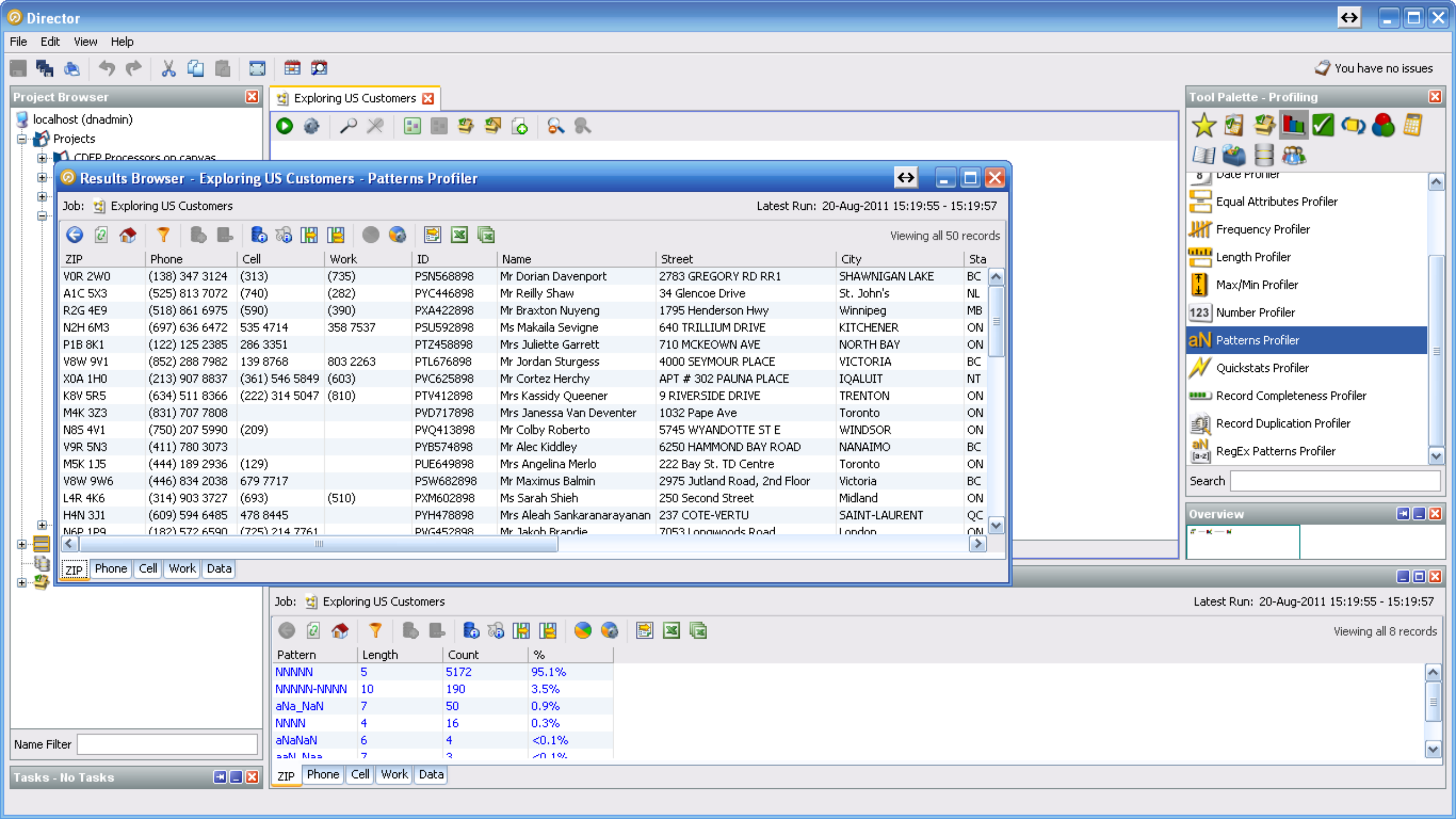

Director
File Edit View Help
You have no issues
Project Browser
localhost (dnadmin)
Projects
Exploring US Customers
Tool Palette - Profiling
Date Profiler
Equal Attributes Profiler
Frequency Profiler
Length Profiler
Max/Min Profiler
Number Profiler
Patterns Profiler
Quickstats Profiler
Record Completeness Profiler
Record Duplication Profiler
RegEx Patterns Profiler
Search
Overview
Results Browser - Exploring US Customers - Quickstats Profiler
Job: Exploring US Customers
Latest Run: 20-Aug-2011 15:12:02 - 15:12:17
Viewing all 4 records
Gender Count %
M 2192 40.3%
F 2153 39.6%
1058 19.5%
U 35 0.6%
Summary statistics view
Data
Latest Run: 20-Aug-2011 15:12:02 - 15:12:17
Viewing all 17 records
Street 5438 5438 0 5319 119 5376 Complete; Potentially damaged key; Investigate duplicates
City 5438 5438 0 396 5042 1232 Complete
State 5438 5438 0 12 5426 65 Complete
Summary statistics view
Data
Name Filter
Tasks - No Tasks

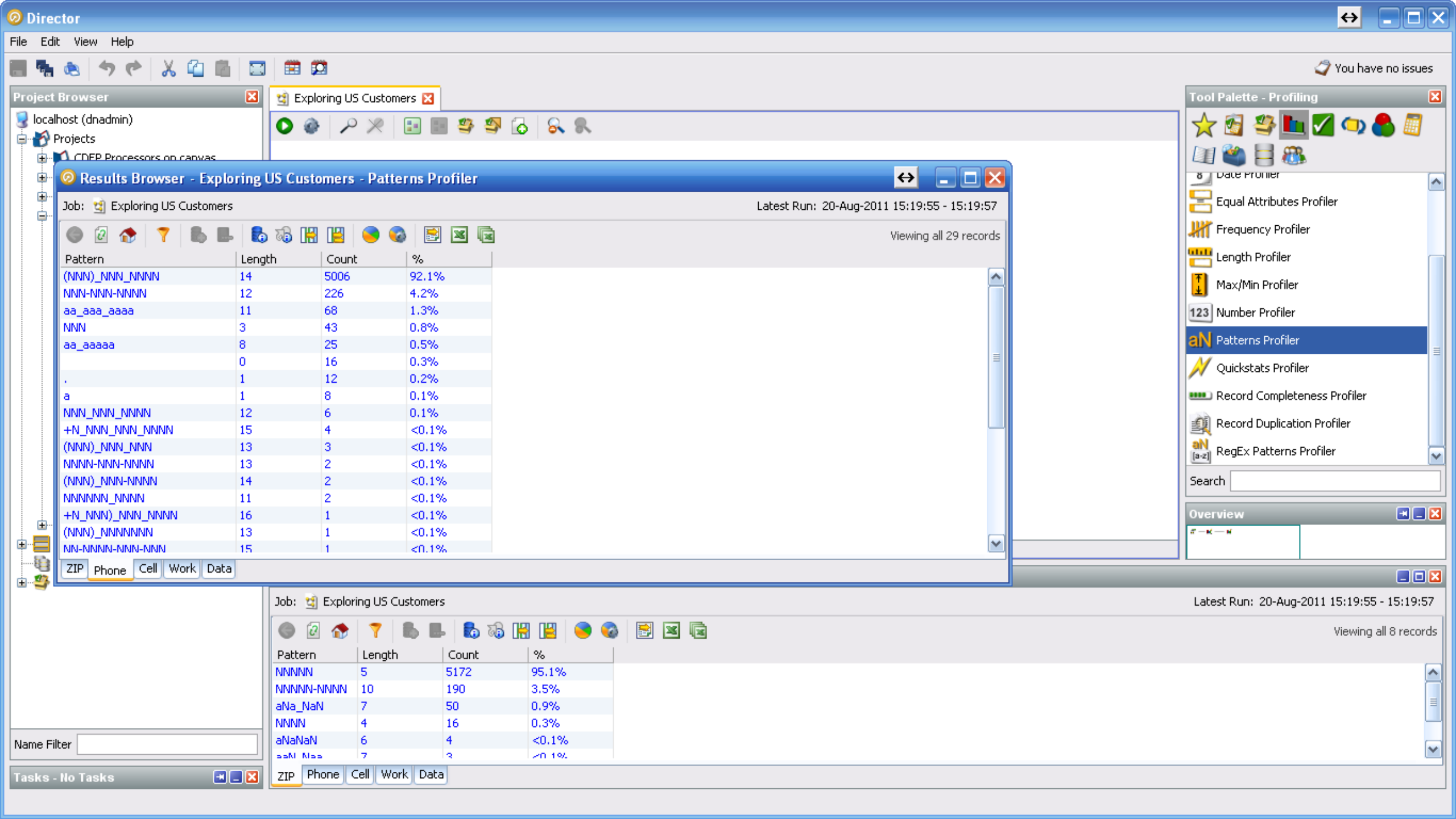

Director
File Edit View Help
You have no issues
Project Browser
localhost (dnadmin)
Projects
Exploring US Customers
Tool Palette - Profiling
Date Profiler
Equal Attributes Profiler
Frequency Profiler
Length Profiler
Max/Min Profiler
Number Profiler
Patterns Profiler
Quickstats Profiler
Record Completeness Profiler
Record Duplication Profiler
RegEx Patterns Profiler
Search
Overview
Results Browser - Exploring US Customers - Quickstats Profiler
Job: Exploring US Customers
Latest Run: 20-Aug-2011 15:12:02 - 15:12:17
Viewing all 35 records
Gender ID Name Street City State ZIP Country Phone
U MBH421308 John Mitchell 380 Beinoris Drive WOOD DALE IL 60191 USA (140) 2
U PUW442949 Irma Bailey 2101 New Beaver Avenue PITTSBURGH PA 15233 USA (121) 8
U VRL714090 Gregory Smith 3932 Ridgeoak Way FARMERS BRANCH TX 75234 (144) 2
U JEH471585 Marion Chan Old River Road MARCY NY 13403 (901) 3
U YJU651180 Amanda Brown 4205 Jasper Court ROWLETT TX 75088 USA (561) 8
U BXS475882 Cecilia Ong & Charles Archer 1854 East Broadway ALTON IL 62002 USA (421) 6
U DRC711036 Victor Mccoy 9300 East Smith Road DENVER CO 80207-1757 (753) 1
U KBS609457 Eileen Person 3015 S Valley Avenue MARION IN 46953 (773) 5
U BSC590670 Rosalind Decker 179 Darnell Lane DRESDEN TN 38225 (165) 3
U PBQ406254 Dr. Jessica Frailey Route 120 and Wilson Road ROUND LAKE IL 60073 U.S.A (629) 5
U FLA554976 Alan Hammond Associates 889 West Johnson Avenue TERRE HAUTE IN 47802 USA (447) 3
U PHF441856 Dr Jeannette Sather 1935 Motor Street DALLAS TX 75235 (294) 5
U HJG532356 Ayesha Holmes 729 Navco Drive LAFAYETTE IN 47905 USA (367) 6
U MZK411179 Mary Collins 1901 N. Fountain Green Road BEL AIR MD 21015-1411 (561) 2
U RSM597707 Tamara Pearson 1 Mirror Lake Drive LAKE PLACID NY 12946 USA 770-29
U NYN525227 Joyce Perry 710 South 9th Street GUNNISON CO 81230 USA (878) 3
U WQG650768 Kenneth Thomas 9009 W Shawnee Mission Parkway MERRIAM KS 66202 USA (153) 3
U QSN402369 Lessie Sanchez 26673 Lawrence CENTER LINE MI 48015 USA (909) 3
U LAZ660348 Kevin Britt 12814 W Denton Avenue LITCHFIELD PARK AZ 85340 USA (651) 3
U HEJ720083 Mr Amado Faison & Ms Pat Butcher 845 Larch Avenue ELMHURST IL 60126 (638) 8
U ANJ609945 Ryan Arter & Emily May 4128 Rockford Road Reno NV 89501 USA 901-79
U AQM457231 Ira Dudley 3348 Honeysuckle Lane Vancouver WA 98686 443-29
Summary statistics view
Data
Latest Run: 20-Aug-2011 15:12:02 - 15:12:17
Viewing all 17 records
Street 5438 5438 0 5319 119 5376 Complete; Potentially damaged key; Investigate duplicates
City 5438 5438 0 396 5042 1232 Complete
State 5438 5438 0 12 5426 65 Complete
Summary statistics view
Data
Name Filter
Tasks - No Tasks

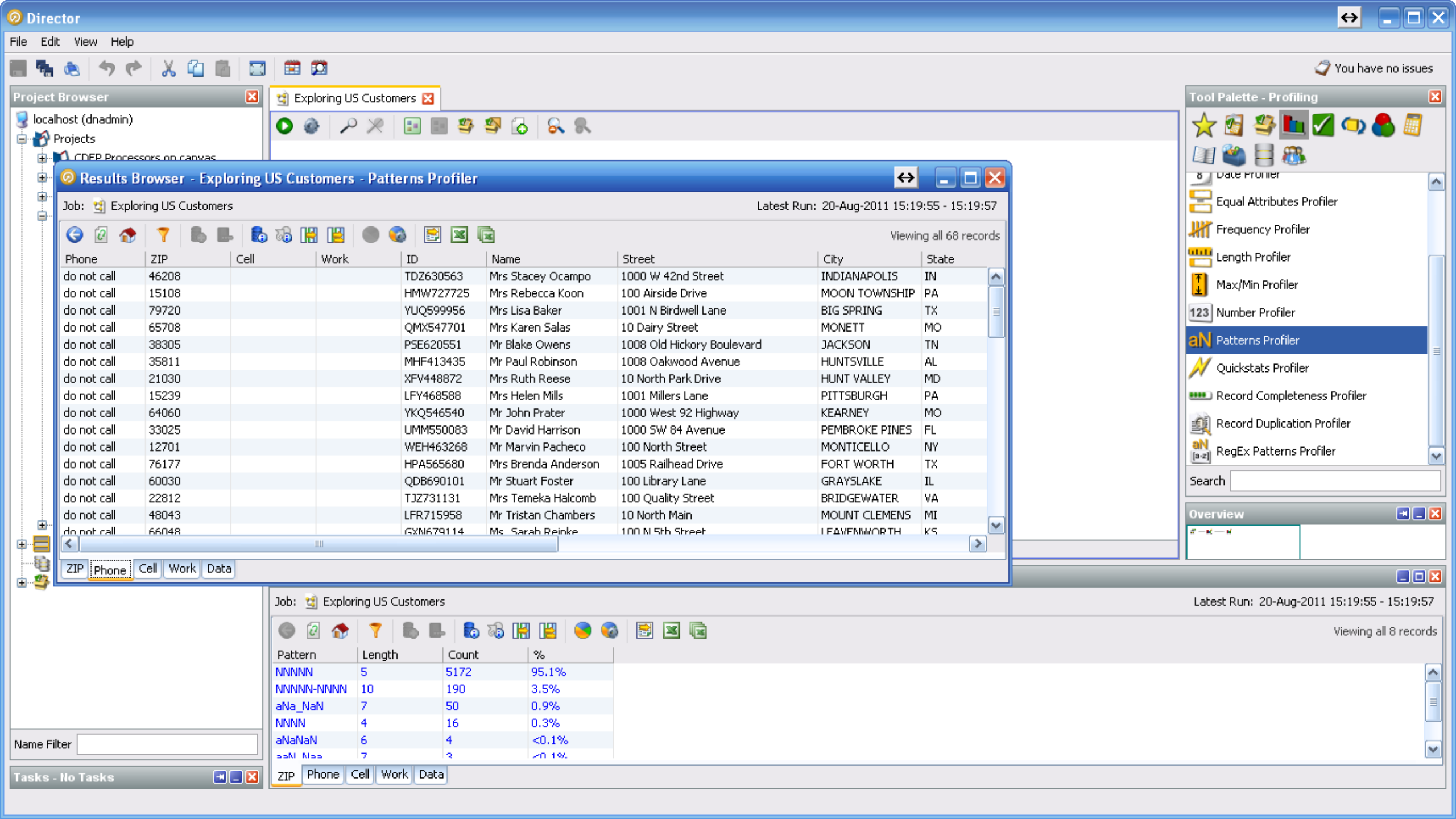

Director
File Edit View Help
You have no issues
Project Browser
localhost (dnadmin)
Projects
CDEP Processors on canvas
Compass Data Quality
Customer data demo - US
Demo
Data Stores
DemoData
Staged Data
US Customers
Views
Processes
Exploring US Customers
Reference Data
Valid Genders
Results Books
External Tasks
Jobs
Exports
Web Services
Notes
Web Service demo
Reference Data
Data Stores
Published Processors
Name Filter
Tasks - No Tasks
Exploring US Customers
Read US Customers
Quickstats Profiler
Patterns Profiler
Process
Tool Palette - Profiling
Equal Attributes Profiler
Frequency Profiler
Length Profiler
Max/Min Profiler
Number Profiler
Patterns Profiler
Quickstats Profiler
Record Completeness Profiler
Record Duplication Profiler
RegEx Patterns Profiler
Search
Overview
Results Browser
Job: Exploring US Customers
Latest Run: 20-Aug-2011 15:12:02 - 15:12:17
This processor did not exist when the process was run.


Director
File Edit View Help
You have no issues
Project Browser
localhost (dnadmin)
Projects
CDEP Processors on canvas
Compass Data Quality
Customer data demo - US
Demo
Data Stores
DemoData
Staged Data
US Customers
Views
Processes
Exploring US Customers
Reference Data
Valid Genders
Results Books
External Tasks
Jobs
Exports
Web Services
Notes
Web Service demo
Reference Data
Data Stores
Published Processors
Name Filter
Tasks - No Tasks
Exploring US Customers
Patterns Profiler
Attributes
Options
Dashboard
Notes
Icon & Family
Available Attributes:
Data
ID (string)
Name (string)
Street (string)
City (string)
State (string)
ZIP (string)
Country (string)
Phone (string)
Cell (string)
Work (string)
eMail (string)
DoB (string)
Gender (string)
Active (string)
CreditLimit (string)
StartDate (string)
EndDate (string)
Added
Flags
All
Processors
Search
Selected Attributes:
ZIP
Phone
Cell
Work
Output Attributes:
Added (0)
Flags (4)
OK
Cancel
Tool Palette - Profiling
Equal Attributes Profiler
Frequency Profiler
Length Profiler
Max/Min Profiler
Number Profiler
Patterns Profiler
Quickstats Profiler
Record Completeness Profiler
Record Duplication Profiler
RegEx Patterns Profiler
Search
Overview
Latest Run: 20-Aug-2011 15:12:02 - 15:12:17
This processor did not exist when the process was run.


Director
File Edit View Help
You have no issues
Project Browser
localhost (dnadmin)
Projects
Exploring US Customers
Results Browser - Exploring US Customers - Patterns Profiler
Job: Exploring US Customers
Latest Run: 20-Aug-2011 15:19:55 - 15:19:57
Viewing all 8 records
Pattern | Length | Count | %
NNNNN | 5 | 5172 | 95.1%
NNNNN-NNNN | 10 | 190 | 3.5%
aNa_NaN | 7 | 50 | 0.9%
NNNN | 4 | 16 | 0.3%
aNaNaN | 6 | 4 | <0.1%
aaN_Naa | 7 | 3 | <0.1%
 | 0 | 2 | <0.1%
aNa-NaN | 7 | 1 | <0.1%
ZIP Phone Cell Work Data
Tool Palette - Profiling
Equal Attributes Profiler
Frequency Profiler
Length Profiler
Max/Min Profiler
Number Profiler
Patterns Profiler
Quickstats Profiler
Record Completeness Profiler
Record Duplication Profiler
RegEx Patterns Profiler
Search
Overview
Name Filter
Tasks - No Tasks


Director
File Edit View Help
You have no issues
Project Browser
localhost (dnadmin)
Projects
Exploring US Customers
Results Browser - Exploring US Customers - Patterns Profiler
Job: Exploring US Customers
Latest Run: 20-Aug-2011 15:19:55 - 15:19:57
Viewing all 50 records
ZIP | Phone | Cell | Work | ID | Name | Street | City | Sta
V0R 2W0 | (138) 347 3124 | (313) | (735) | PSN568898 | Mr Dorian Davenport | 2783 GREGORY RD RR1 | SHAWNIGAN LAKE | BC
A1C 5X3 | (525) 813 7072 | (740) | (282) | PYC446898 | Mr Reilly Shaw | 34 Glencoe Drive | St. John's | NL
R2G 4E9 | (518) 861 6975 | (590) | (390) | PXA422898 | Mr Braxton Nuyeng | 1795 Henderson Hwy | Winnipeg | MB
N2H 6M3 | (697) 636 6472 | 535 4714 | 358 7537 | PSU592898 | Ms Makaila Sevigne | 640 TRILLIUM DRIVE | KITCHENER | ON
P1B 8K1 | (122) 125 2385 | 286 3351 | | PTZ458898 | Mrs Juliette Garrett | 710 MCKEOWN AVE | NORTH BAY | ON
V8W 9V1 | (852) 288 7982 | 139 8768 | 803 2263 | PTL676898 | Mr Jordan Sturgess | 4000 SEYMOUR PLACE | VICTORIA | BC
X0A 1H0 | (213) 907 8837 | (361) 546 5849 | (603) | PVC625898 | Mr Cortez Herchy | APT # 302 PAUNA PLACE | IQALUIT | NT
K8V 5R5 | (634) 511 8366 | (222) 314 5047 | (810) | PTV412898 | Mrs Kassidy Queener | 9 RIVERSIDE DRIVE | TRENTON | ON
M4K 3Z3 | (831) 707 7808 | | | PVD717898 | Mrs Janessa Van Deventer | 1032 Pape Ave | Toronto | ON
N8S 4V1 | (750) 207 5990 | (209) | | PVQ413898 | Mr Colby Roberto | 5745 WYANDOTTE ST E | WINDSOR | ON
V9R 5N3 | (411) 780 3073 | | | PYB574898 | Mr Alec Kiddley | 6250 HAMMOND BAY ROAD | NANAIMO | BC
M5K 1J5 | (444) 189 2936 | (129) | | PUE649898 | Mrs Angelina Merlo | 222 Bay St. TD Centre | Toronto | ON
V8W 9W6 | (446) 834 2038 | 679 7717 | | PSW682898 | Mr Maximus Balmin | 2975 Jutland Road, 2nd Floor | Victoria | BC
L4R 4K6 | (314) 903 3727 | (693) | (510) | PXM602898 | Ms Sarah Shieh | 250 Second Street | Midland | ON
H4N 3J1 | (609) 594 6485 | 478 8445 | | PYH478898 | Mrs Aleah Sankaranarayanan | 237 COTE-VERTU | SAINT-LAURENT | QC
ZIP Phone Cell Work Data
Tool Palette - Profiling
Equal Attributes Profiler
Frequency Profiler
Length Profiler
Max/Min Profiler
Number Profiler
Patterns Profiler
Quickstats Profiler
Record Completeness Profiler
Record Duplication Profiler
RegEx Patterns Profiler
Search
Overview
Job: Exploring US Customers
Latest Run: 20-Aug-2011 15:19:55 - 15:19:57
Viewing all 8 records
Pattern | Length | Count | %
NNNNN | 5 | 5172 | 95.1%
NNNNN-NNNN | 10 | 190 | 3.5%
aNa_NaN | 7 | 50 | 0.9%
NNNN | 4 | 16 | 0.3%
aNaNaN | 6 | 4 | <0.1%
ZIP Phone Cell Work Data
Name Filter
Tasks - No Tasks


Director
File Edit View Help
You have no issues
Project Browser
localhost (dnadmin)
Projects
Exploring US Customers
Tool Palette - Profiling
Date Profiler
Equal Attributes Profiler
Frequency Profiler
Length Profiler
Max/Min Profiler
Number Profiler
Patterns Profiler
Quickstats Profiler
Record Completeness Profiler
Record Duplication Profiler
RegEx Patterns Profiler
Search
Overview
Results Browser - Exploring US Customers - Patterns Profiler
Job: Exploring US Customers
Latest Run: 20-Aug-2011 15:19:55 - 15:19:57
Viewing all 29 records
Pattern | Length | Count | %
(NNN)_NNN_NNNN | 14 | 5006 | 92.1%
NNN-NNN-NNNN | 12 | 226 | 4.2%
aa_aaa_aaaa | 11 | 68 | 1.3%
NNN | 3 | 43 | 0.8%
aa_aaaaa | 8 | 25 | 0.5%
 | 0 | 16 | 0.3%
. | 1 | 12 | 0.2%
a | 1 | 8 | 0.1%
NNN_NNN_NNNN | 12 | 6 | 0.1%
+N_NNN_NNN_NNNN | 15 | 4 | <0.1%
(NNN)_NNN_NNN | 13 | 3 | <0.1%
NNNN-NNN-NNNN | 13 | 2 | <0.1%
(NNN)_NNN-NNNN | 14 | 2 | <0.1%
NNNNNN_NNNN | 11 | 2 | <0.1%
+N_NNN)_NNN_NNNN | 16 | 1 | <0.1%
(NNN)_NNNNNNN | 13 | 1 | <0.1%
ZIP Phone Cell Work Data
Job: Exploring US Customers
Latest Run: 20-Aug-2011 15:19:55 - 15:19:57
Viewing all 8 records
Pattern | Length | Count | %
NNNNN | 5 | 5172 | 95.1%
NNNNN-NNNN | 10 | 190 | 3.5%
aNa_NaN | 7 | 50 | 0.9%
NNNN | 4 | 16 | 0.3%
aNaNaN | 6 | 4 | <0.1%
ZIP Phone Cell Work Data
Name Filter
Tasks - No Tasks


Director
File Edit View Help
You have no issues
Project Browser
localhost (dnadmin)
Projects
Exploring US Customers
Tool Palette - Profiling
Date Profiler
Equal Attributes Profiler
Frequency Profiler
Length Profiler
Max/Min Profiler
Number Profiler
Patterns Profiler
Quickstats Profiler
Record Completeness Profiler
Record Duplication Profiler
RegEx Patterns Profiler
Search
Overview
Results Browser - Exploring US Customers - Patterns Profiler
Job: Exploring US Customers
Latest Run: 20-Aug-2011 15:19:55 - 15:19:57
Viewing all 68 records
Phone ZIP Cell Work ID Name Street City State
do not call 46208 TDZ630563 Mrs Stacey Ocampo 1000 W 42nd Street INDIANAPOLIS IN
do not call 15108 HMW727725 Mrs Rebecca Koon 100 Airside Drive MOON TOWNSHIP PA
do not call 79720 YUQ599956 Mrs Lisa Baker 1001 N Birdwell Lane BIG SPRING TX
do not call 65708 QMX547701 Mrs Karen Salas 10 Dairy Street MONETT MO
do not call 38305 PSE620551 Mr Blake Owens 1008 Old Hickory Boulevard JACKSON TN
do not call 35811 MHF413435 Mr Paul Robinson 1008 Oakwood Avenue HUNTSVILLE AL
do not call 21030 XFV448872 Mrs Ruth Reese 10 North Park Drive HUNT VALLEY MD
do not call 15239 LFY468588 Mrs Helen Mills 1001 Millers Lane PITTSBURGH PA
do not call 64060 YKQ546540 Mr John Prater 1000 West 92 Highway KEARNEY MO
do not call 33025 UMM550083 Mr David Harrison 1000 SW 84 Avenue PEMBROKE PINES FL
do not call 12701 WEH463268 Mr Marvin Pacheco 100 North Street MONTICELLO NY
do not call 76177 HPA565680 Mrs Brenda Anderson 1005 Railhead Drive FORT WORTH TX
do not call 60030 QDB690101 Mr Stuart Foster 100 Library Lane GRAYSLAKE IL
do not call 22812 TJZ731131 Mrs Temeka Halcomb 100 Quality Street BRIDGEWATER VA
do not call 48043 LFR715958 Mr Tristan Chambers 10 North Main MOUNT CLEMENS MI
ZIP Phone Cell Work Data
Job: Exploring US Customers
Latest Run: 20-Aug-2011 15:19:55 - 15:19:57
Viewing all 8 records
Pattern Length Count %
NNNNN 5 5172 95.1%
NNNNN-NNNN 10 190 3.5%
aNa_NaN 7 50 0.9%
NNNN 4 16 0.3%
aNaNaN 6 4 <0.1%
ZIP Phone Cell Work Data
Name Filter
Tasks - No Tasks

Director

File Edit View Help

You have no issues

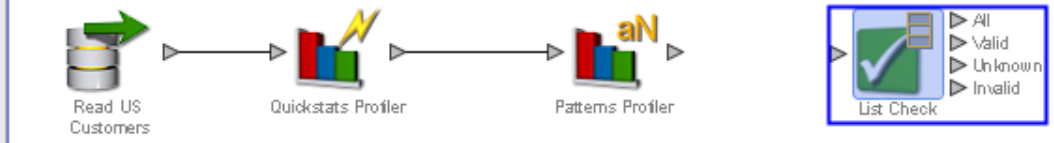
Project Browser

- localhost (dnadmin)
  - Projects
    - CDEP Processors on canvas
    - Compass Data Quality
    - Customer data demo - US
    - Demo
      - Data Stores
        - DemoData
      - Staged Data
        - US Customers
      - Views
      - Processes
        - Exploring US Customers
      - Reference Data
        - Valid Genders
      - Results Books
      - External Tasks
      - Jobs
      - Exports
      - Web Services
      - Notes
    - Web Service demo
  - Reference Data
  - Data Stores
  - Published Processors

Name Filter

Tasks - No Tasks

Exploring US Customers



Process

Tool Palette - Audit

- Business Rules Check
- Cross Attribute Check
- Data Type Check
- Duplicate Check
- Email Check
- GBR Postcode Format Check
- Invalid Character Check
- Length Check
- List Check
- Logic Check
- Lookup Check

Search

Overview

Results Browser

Job: Exploring US Customers

Latest Run: 20-Aug-2011 15:19:55 - 15:19:57

This processor did not exist when the process was run.

Director

File Edit View Help

You have no issues

Project Browser

- localhost (dnadmin)
- Projects
  - CDEP Processors on canvas
  - Compass Data Quality
  - Customer data demo - US
  - Demo
    - Data Stores
      - DemoData
    - Staged Data
      - US Customers
    - Views
    - Processes
      - Exploring US Customers
    - Reference Data
      - Valid Genders
    - Results Books
    - External Tasks
    - Jobs
    - Exports
    - Web Services
    - Notes
  - Web Service demo
- Reference Data
- Data Stores
- Published Processors

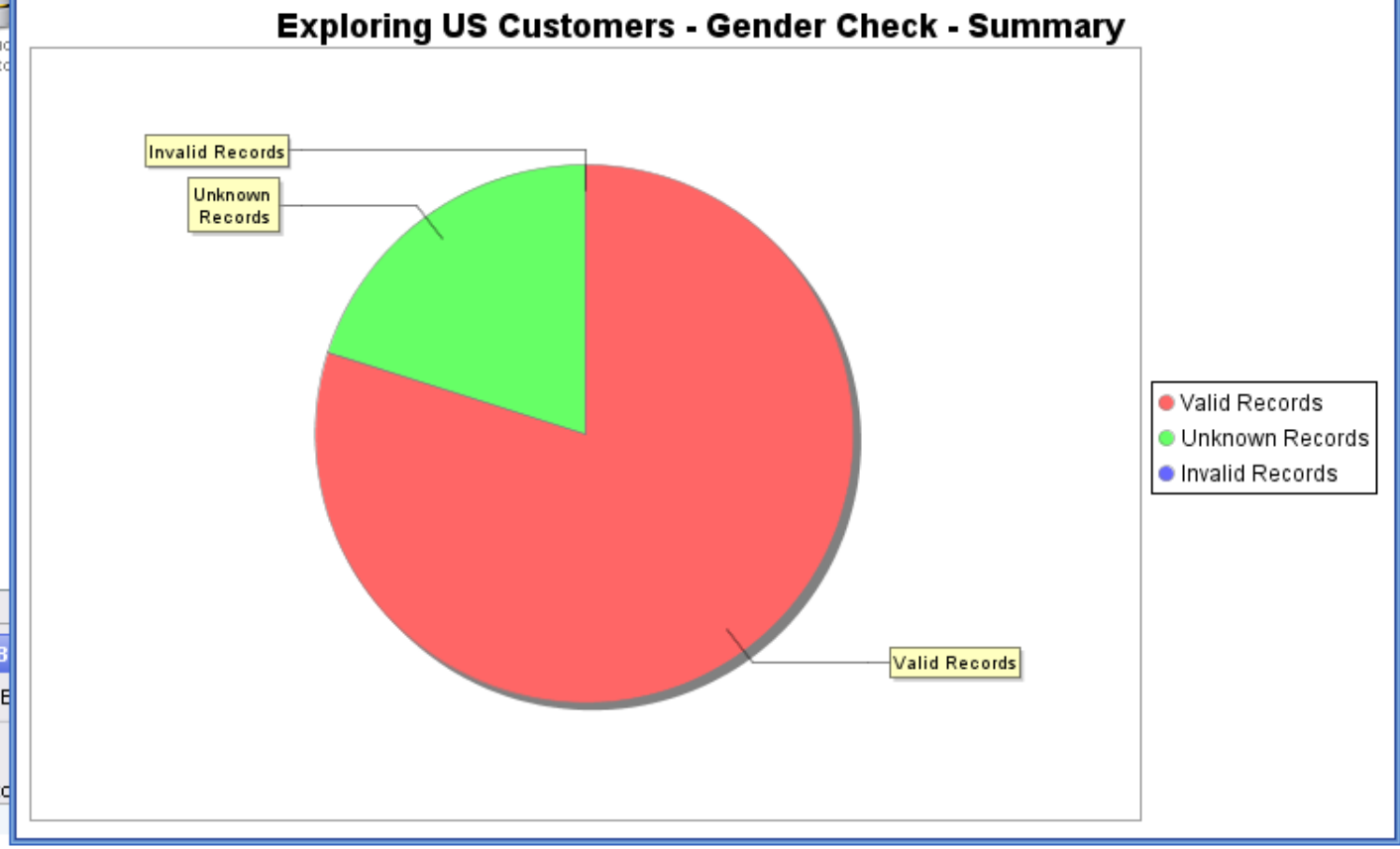
Name Filter

Tasks - No Tasks

Exploring US Customers

Process

Exploring US Customers - Gender Check - Summary



Results B

Job:

Valid Recc

4345

Summary Data

Tool Palette - Audit

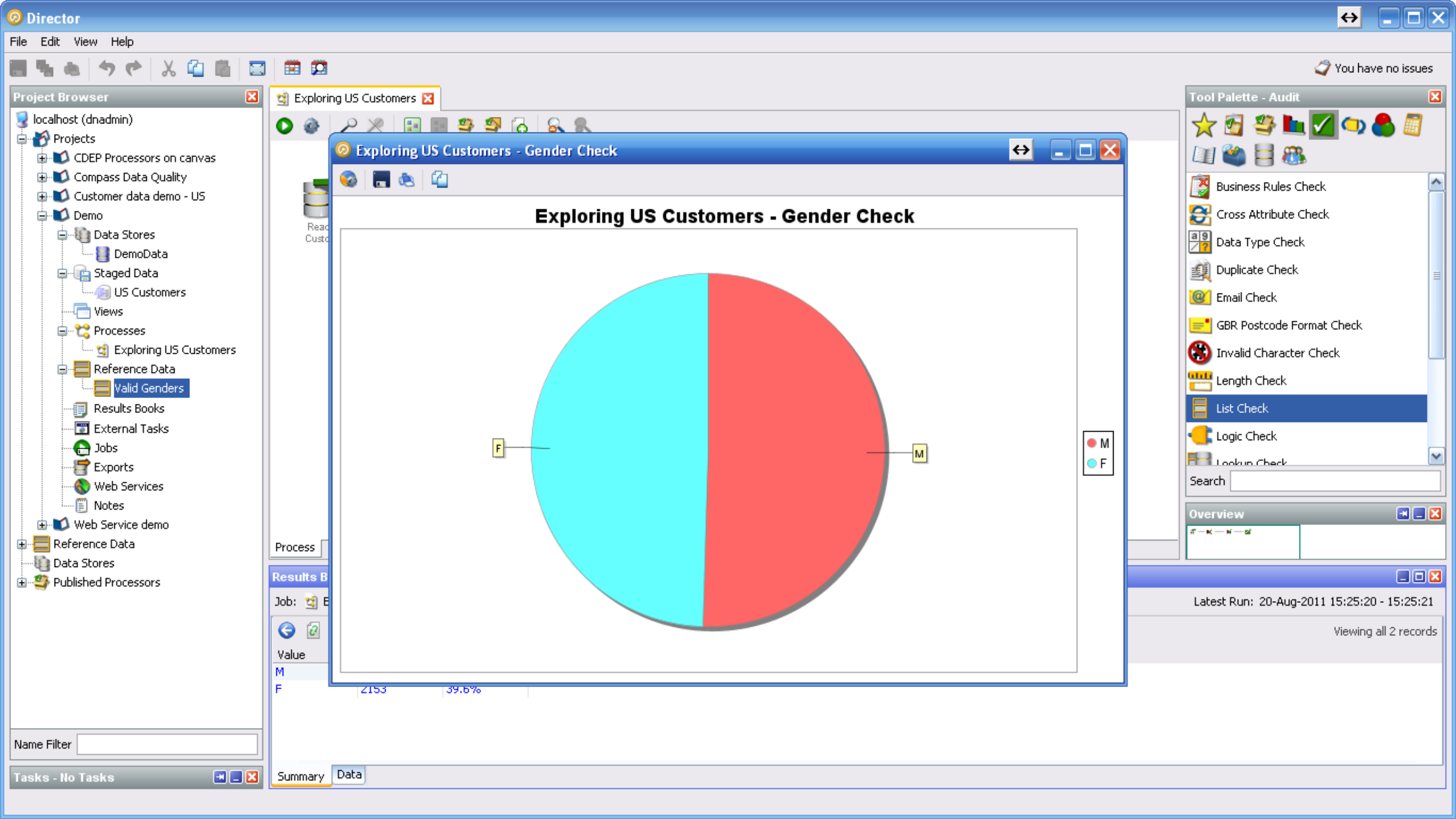
- Business Rules Check
- Cross Attribute Check
- Data Type Check
- Duplicate Check
- Email Check
- GBR Postcode Format Check
- Invalid Character Check
- Length Check
- List Check
- Logic Check
- Lookup Check

Search

Overview

Latest Run: 20-Aug-2011 15:25:20 - 15:25:21

Viewing all 1 records


Director
File Edit View Help
You have no issues
Project Browser
localhost (dnadmin)
Projects
CDEP Processors on canvas
Compass Data Quality
Customer data demo - US
Demo
Data Stores
DemoData
Staged Data
US Customers
Views
Processes
Exploring US Customers
Reference Data
Valid Genders
Results Books
External Tasks
Jobs
Exports
Web Services
Notes
Web Service demo
Reference Data
Data Stores
Published Processors
Name Filter
Tasks - No Tasks
Exploring US Customers
Exploring US Customers - Gender Check
Exploring US Customers - Gender Check
F
M
M
F
Process
Results B
Job:
Value
M
F
2153
39.6%
Summary
Data
Latest Run: 20-Aug-2011 15:25:20 - 15:25:21
Viewing all 2 records
Tool Palette - Audit
Business Rules Check
Cross Attribute Check
Data Type Check
Duplicate Check
Email Check
GBR Postcode Format Check
Invalid Character Check
Length Check
List Check
Logic Check
Search
Overview

Director

File Edit View Help

You have no issues

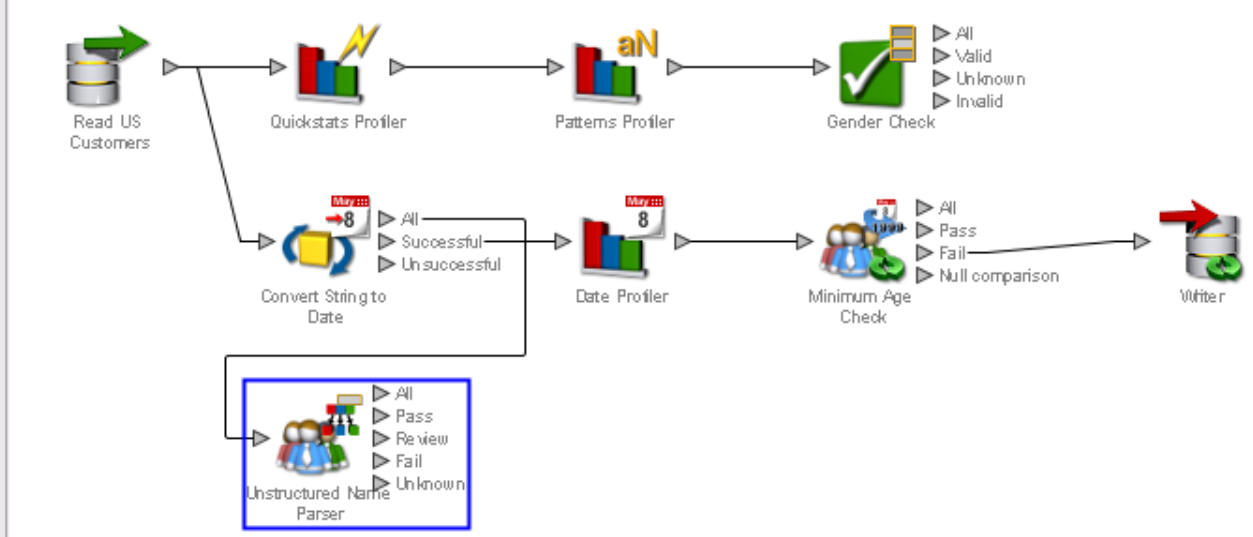
Project Browser

- localhost (dnadmin)
- Projects
  - CDEP Processors on canvas
  - Compass Data Quality
  - Customer data demo - US
  - Demo
    - Data Stores
      - DemoData
    - Staged Data
      - Underage Customers
      - US Customers
    - Views
    - Processes
      - Exploring US Customers
    - Reference Data
      - Valid Genders
    - Results Books
    - External Tasks
    - Jobs
    - Exports
    - Web Services
    - Notes
  - Web Service demo
- Reference Data
- Data Stores
- Published Processors

Name Filter

Tasks - No Tasks

Exploring US Customers



Process

Tool Palette - Customer Data

- GeoNames Country Codes from City
- Get Year from Date
- Match Entities
- Match Households
- Match Individuals (Name, Address, DoB)
- Match Individuals (Name, Address)
- Profile Entity Names
- Standardize Country Names
- Standardize Entity Names
- Structured Name Parser

Search

Overview

Results Browser

Job: Exploring US Customers

Latest Run: 20-Aug-2011 15:52:11 - 15:52:45

Viewing 100 records of 5,438

Name	UnclassifiedData	P1_Prefix	P1_First	P1_Middle	P1_Last	P1_Suffix	A1_Prefix	A1_First	A1_Middle	A1_Last
Mr Everett Hughes & Miss Jane Young	&	Mr	Everett	Hughes	Young					
Mr Edward Yates & Mrs Evangelina Yates	&	Mr	Edward	Yates	Yates					
Mrs Maria Verdin and Dr John Roberts	and	Mrs	Maria	Verdin	Roberts					
Ryan Arter & Emily May	&		Ryan		Arter					

Base Tokenization | Token Checks | Classification | Unclassified Tokens | Reclassification Rules | Reclassification | Selection | Resolution Rules | Results | Data

Director

File Edit View Help

You have no issues

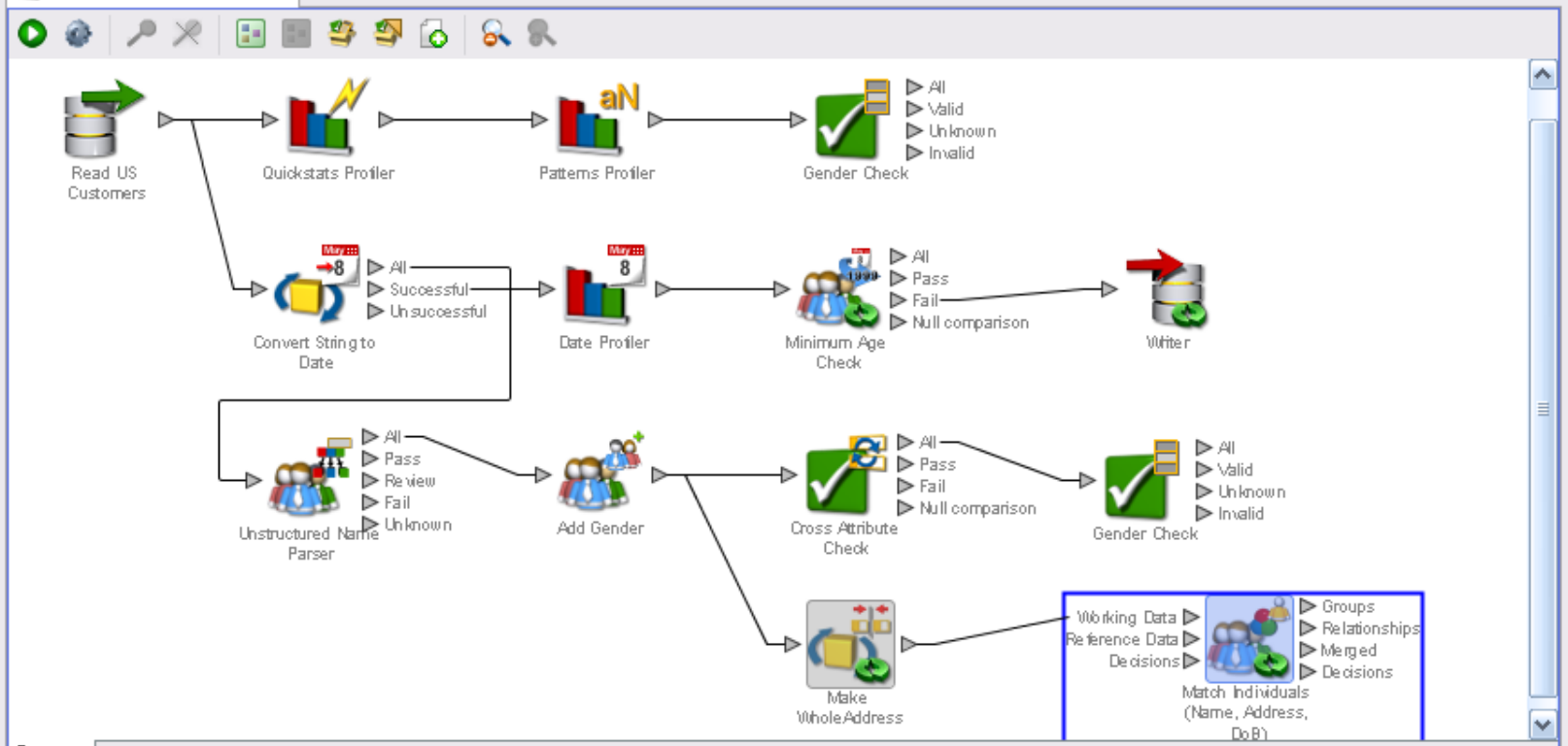
Project Browser

- localhost (dnadmin)
- Projects
  - CDEP Processors on canvas
  - Compass Data Quality
  - Customer data demo - US
  - Demo
    - Data Stores
      - DemoData
    - Staged Data
      - Underage Customers
      - US Customers
    - Views
    - Processes
      - Exploring US Customers
    - Reference Data
      - Valid Genders
    - Results Books
      - Gender improvement
    - External Tasks
    - Jobs
    - Exports
    - Web Services
    - Notes
  - Web Service demo
- Reference Data
- Data Stores
- Published Processors

Name Filter

Tasks - No Tasks

Exploring US Customers



Process

Tool Palette - Customer Data

- Match Entities
- Match Households
- Match Individuals (Name, Address, DoB)
- Match Individuals (Name, Address)
- Profile Entity Names
- Standardize Country Names
- Standardize Entity Names
- Structured Name Parser
- Unstructured Name Parser
- URL Check

Search

Overview

Results Browser

Job: Exploring US Customers

Latest Run: 20-Aug-2011 16:02:34 - 16:02:50

This processor did not exist when the process was run.

Director

File Edit View Help

You have no issues

Project Browser

- localhost (dnadmin)
  - Projects
    - CDEP Processors on canvas
    - Compass Data Quality
    - Customer data demo - US
    - Demo
      - Data Stores
        - DemoData
      - Staged Data
        - Underage Customers
        - US Customers
      - Views
      - Processes
        - Exploring US Customers
      - Reference Data
        - Valid Genders
      - Results Books
        - Gender improvement
      - External Tasks
      - Jobs
      - Exports
      - Web Services
      - Notes
    - Web Service demo
  - Reference Data
  - Data Stores
  - Published Processors

Name Filter

Tasks - No Tasks

Exploring US Customers

Match Individuals (Name, Address, DoB)

Advanced Options | Assign Relationship Review
Review Results | Assign Merged Review
Configure Bulk Review Rules | View Match Statistics
Delete Realtime Review Results | Delete Manual Decisions

Input Identify Cluster Match Merge

Rule	Priority	Decision
DOB exact, Name, Address	0	MATCH
DOB no data, Name, Address	0	MATCH
DOB exact, Name, Postcode	0	MATCH
DOB transposed, Name, Address	0	MATCH
DOB close, Name, Address	0	MATCH
YOB, Name, Address	0	MATCH
DOB exact, Name typos, Address	0	MATCH

Show Summary

Process | Match Individuals (Name, Address, DoB)

Tool Palette - Customer Data

- GeoNames Country Codes from City
- Get Year from Date
- Match Entities
- Match Households
- Match Individuals (Name, Address, DoB)
- Match Individuals (Name, Address)
- Profile Entity Names
- Standardize Country Names
- Standardize Entity Names
- Structured Name Parser
- Unstructured Name Parser
- URL Check

Search

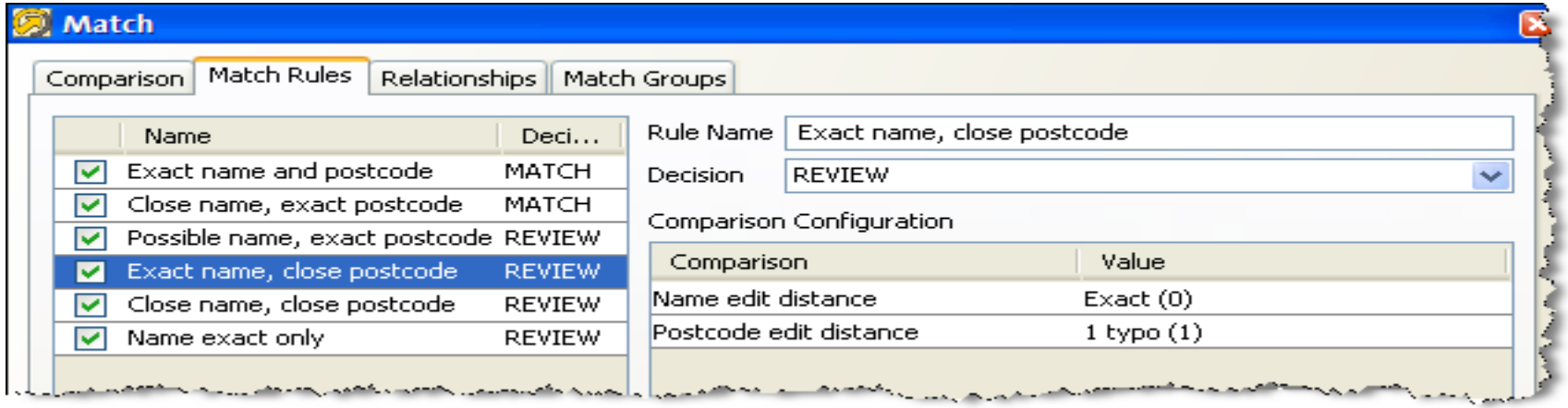
Overview

Results Browser

Job: Exploring US Customers

Latest Run: 20-Aug-2011 16:02:34 - 16:02:50

This processor did not exist when the process was run.



- A Match Rule is simply the combination of comparison results
- Rules are evaluated in order and if one hits, we stop
- Rules can be ‘negative’ to eliminate pairs that are too different with a ‘No Match’ rule
- Rules can easily be turned on & off during the tuning process

Director

File Edit View Help

You have no issues

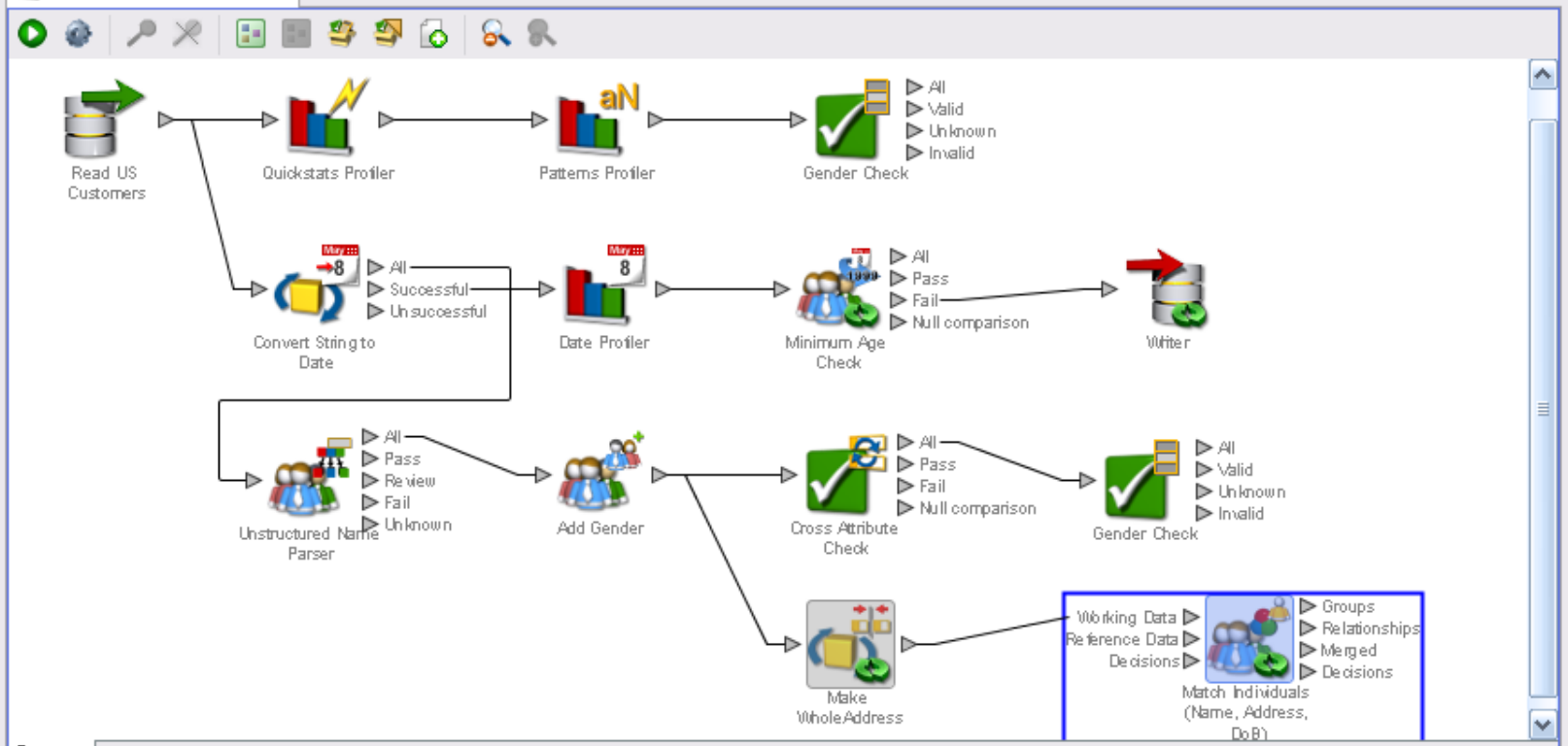
Project Browser

- localhost (dnadmin)
- Projects
  - CDEP Processors on canvas
  - Compass Data Quality
  - Customer data demo - US
  - Demo
    - Data Stores
      - DemoData
    - Staged Data
      - Underage Customers
      - US Customers
    - Views
    - Processes
      - Exploring US Customers
    - Reference Data
      - Valid Genders
    - Results Books
      - Gender improvement
    - External Tasks
    - Jobs
    - Exports
    - Web Services
    - Notes
  - Web Service demo
- Reference Data
- Data Stores
- Published Processors

Name Filter

Tasks - No Tasks

Exploring US Customers



Process

Tool Palette - Customer Data

- Match Entities
- Match Households
- Match Individuals (Name, Address, DoB)
- Match Individuals (Name, Address)
- Profile Entity Names
- Standardize Country Names
- Standardize Entity Names
- Structured Name Parser
- Unstructured Name Parser
- URL Check

Search

Overview

Results Browser

Job: Exploring US Customers

Latest Run: 20-Aug-2011 16:02:34 - 16:02:50

This processor did not exist when the process was run.

Director

File Edit View Help

You have no issues

Project Browser

- localhost (dnadmin)
  - Projects
    - CDEP Processors on canvas
    - Compass Data Quality
    - Customer data demo - US
    - Demo
      - Data Stores
        - DemoData
      - Staged Data
        - Underage Customers
        - US Customers
      - Views
      - Processes
        - Exploring US Customers
      - Reference Data
        - Valid Genders
      - Results Books
        - Gender improvement
      - External Tasks
      - Jobs
      - Exports
      - Web Services
      - Notes
    - Web Service demo
  - Reference Data
  - Data Stores
  - Published Processors

Name Filter

Tasks - No Tasks

Matching Records Review - Demo\Exploring US Customers\Match Individuals (Name, Address, DoB)

Viewing group 1 of 8 | Additional groups are available for loading | Review merged output

Review | Review Merged Output

Filter Groups

Look For | Search In | Find | Use OR logic | Case Sensitive | Exact Match | Clear

Review status = Awaiting Review

Records — Highlight Differences

	Internal ID	Match Group	Data Stream	Title	Given Names	Family Name	Full Name	DOB	YOB	Address1
R1	44	62	US Customers	Ms	Sheila	Gibson	Ms. Sheila Gibson	14-Nov-1959 00:00:00		11001 West 120th Aven
R2	1242	1242	US Customers	Ms	Sheila	Gibson	Ms. Sheila Gibson	14-Nov-1959 00:00:00		111 Emerson Street

Relationships — Auto select relationships based on search

Record 1	Group 1	Record 2	Group 2		Match Rule	Review Status	Decision	User	Date
R1	62	R2	1242	*	DOB exact, Name	Awaiting Review	Possible Match		

Tool Palette - Customer Data

- GeoNames Country Codes from City
- Get Year from Date
- Match Entities
- Match Households
- Match Individuals (Name, Address, DoB)
- Match Individuals (Name, Address)
- Profile Entity Names
- Standardize Country Names
- Standardize Entity Names
- Structured Name Parser
- Unstructured Name Parser
- URL Check

Search

Overview

Latest Run: 20-Aug-2011 16:10:55 - 16:11:29

Viewing all 27 records

Data Stream: US Customers | First 4 surname meta + first 4 postcode | First 2 forename stand + YOB + First 2 Postcode + First 3 surname | Matching | Rules

Review Status | Match Groups | Merge Summary | Groups Output | Relationships Output | Merged Output

Director

File Edi

Project

local

issues

DoB)

:11:29

records

Rules

Name Fil

Tasks -

Matching Records Review - Demo\Exploring US Customers\Match Individuals (Name, Address, DoB)

Viewing group 1 of 8

Additional groups are available for loading

Review merged output

Review | Review Merged Output

Filter Groups

Look For

Search In

Find

Use OR logic

Case Sensitive

Exact Match

Clear

Match rule name | = | DOB exact, Name, Address (8)

Records

Highlight Differences

	Internal ID	Match Group	Data Stream	Title	Given Names	Family Name	Full Name	DOB	YOB	Address1	Postcode	WholeAddress	Email
R1	82	1	US Customers	Mrs	Christine	Hunt	Mrs Christine Hunt	06-May-1985 00:00:00		11197 Leadbetter Road	23005	11197 Leadbetter Road ASHLAND VA 23005	Christine.G.Hunt@excite.co
R2	5417	1	US Customers	Mrs	Chris	Hunt	Mrs Chris Hunt	06-May-1985 00:00:00		11197 Leadbetter Rd	23005	11197 Leadbetter Rd ASHLAND VA 23005	Christine.G.Hunt@excite.co

Relationships

Auto select relationships based on search

Record 1	Group 1	Record 2	Group 2		Match Rule	Review Status	Decision	User	Date
R1	1	R2	1	*	DOB exact, Name, Address	No Review Required	Match		

Why EDQ?

- Let's try the 'brute force' approach to **de-duplicating 10 million records**:
- Start at record 1 and compare it with:
  - Record 2, Record 3, … , Record 10 million
- Now move onto record 2 and compare it with:
  - Record 3, Record 4, … , Record 10 million

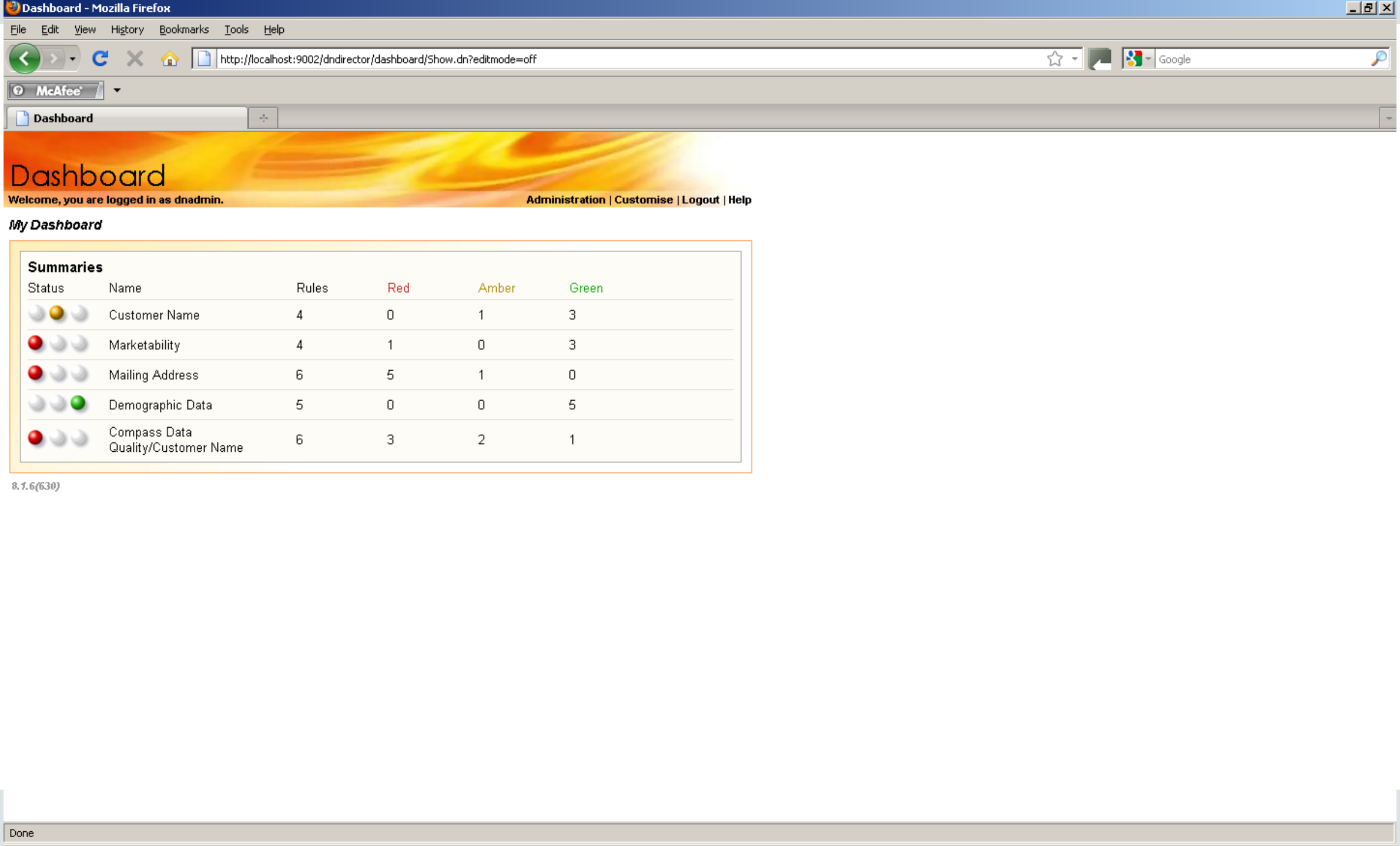
- This could take some time…
- The number of comparisons is about
  - Half of 10 million x 10 million, which is
  - 50,000,000,000,000!
- If a **server can do 100,000 per second** it will take
  - 500,000,000 seconds OR
  - 138889 hours OR
  - **15.85 years**
- Which is rather too long to wait!
- So we need to work a bit smarter…

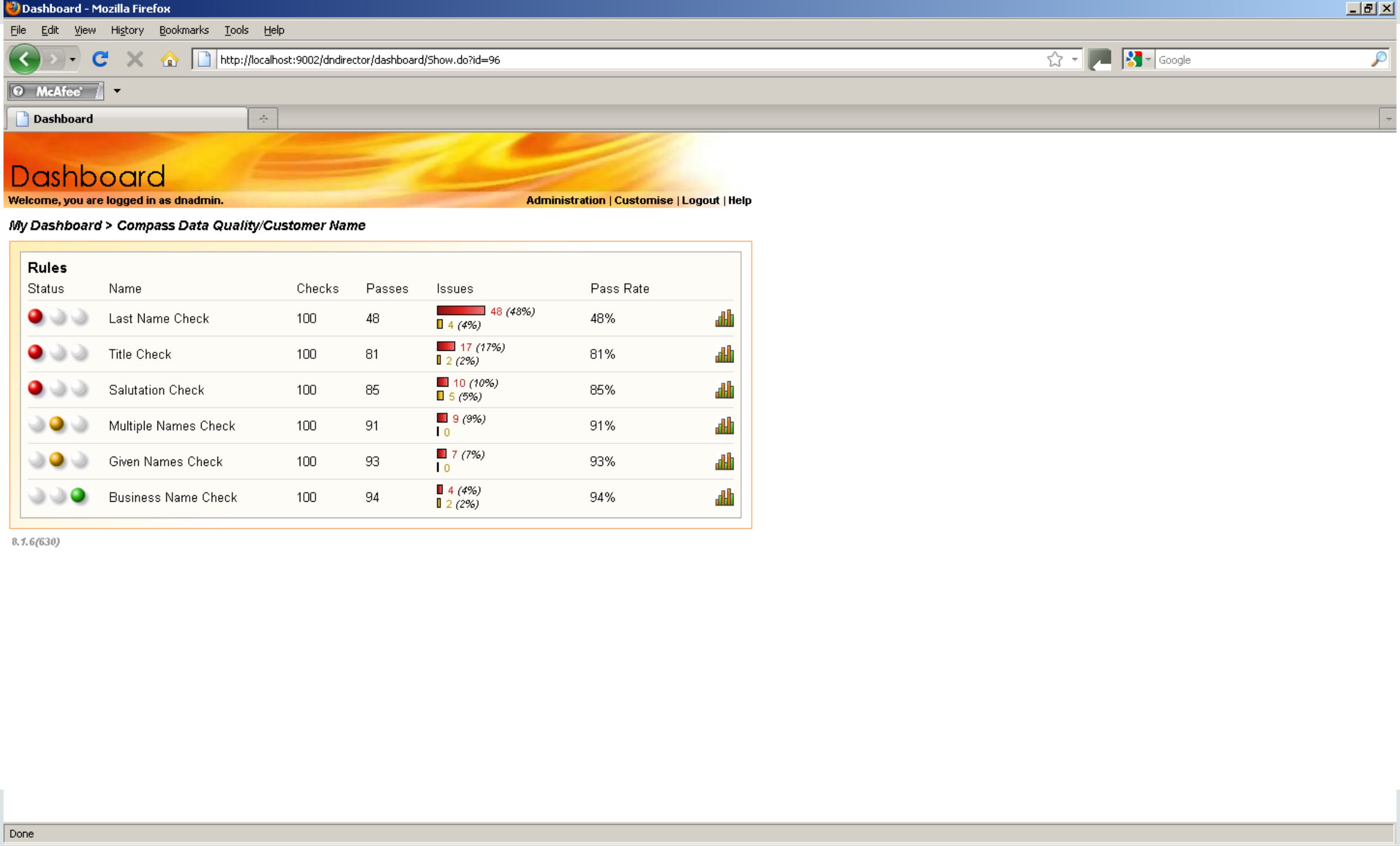
EDQ Benchmark

Profile, Audit, Standardize, and Match 20m customer records in less than 7 hours on a single 6 core (3 pricing processor) server*

Benchmark Test	Time	Rate
Data Capture	9min	35,563 records/second
Profiling	1 hr, 30 min	3,536 records/second
Audit & Dashboard	7min	45,977 records/second
Cleanse	20min	16,474 records/second
Address Verification	5 hrs, 15min	1,059 records/second
Match	4 hrs, 27min	1,176 records/second
Siebel Batch Match	6 hrs, 59min	796 records/second

- 20million records (20 files of 1million records each)
- 6 core Intel/Windows box (3 processors for licensing purposes)


Dashboard - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://localhost:9002/dndirector/dashboard/Show.dn?editmode=off
Google
McAfee
Dashboard
Dashboard
Welcome, you are logged in as dnadmin.
Administration | Customise | Logout | Help
My Dashboard
Summaries
Status Name Rules Red Amber Green
Customer Name 4 0 1 3
Marketability 4 1 0 3
Mailing Address 6 5 1 0
Demographic Data 5 0 0 5
Compass Data Quality/Customer Name 6 3 2 1
8.1.6(630)
Done


Dashboard - Mozilla Firefox
File Edit View History Bookmarks Tools Help
http://localhost:9002/dndirector/dashboard/Show.do?id=96
Google
McAfee
Dashboard
Dashboard
Welcome, you are logged in as dnadmin.
Administration | Customise | Logout | Help
My Dashboard > Compass Data Quality/Customer Name
Rules
Status Name Checks Passes Issues Pass Rate
Last Name Check 100 48 48 (48%) 4 (4%) 48%
Title Check 100 81 17 (17%) 2 (2%) 81%
Salutation Check 100 85 10 (10%) 5 (5%) 85%
Multiple Names Check 100 91 9 (9%) 0 91%
Given Names Check 100 93 7 (7%) 0 93%
Business Name Check 100 94 4 (4%) 2 (2%) 94%
8.1.6(630)
Done

Dashboard - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://localhost:9002/dndirector/dashboard/History.do?id=98

Google

McAfee

Dashboard

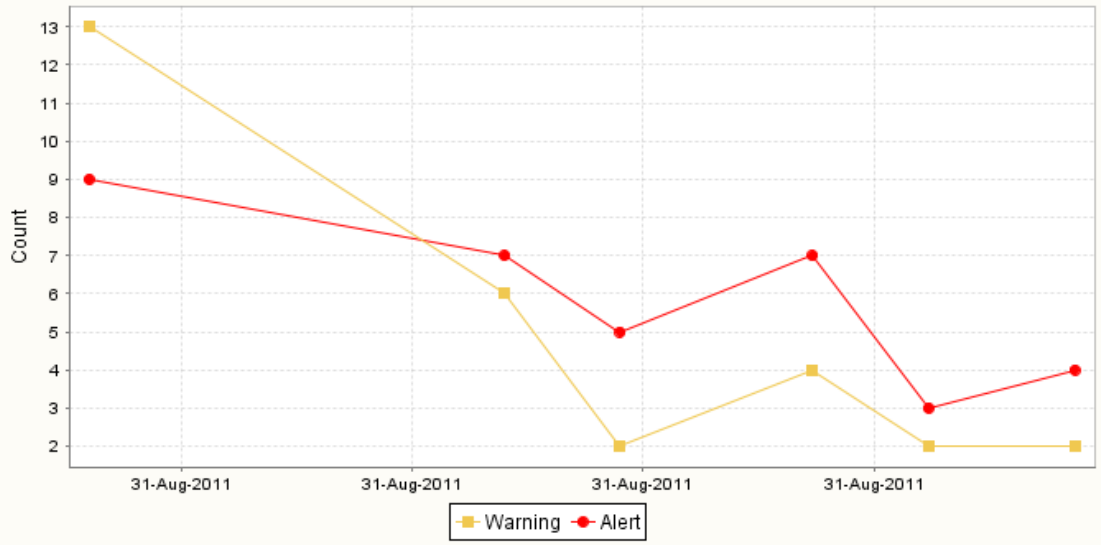
Dashboard

Welcome, you are logged in as dnadmin. Administration | Logout | Help

My Dashboard > Compass Data Quality/Customer Name > Business Name Check

History

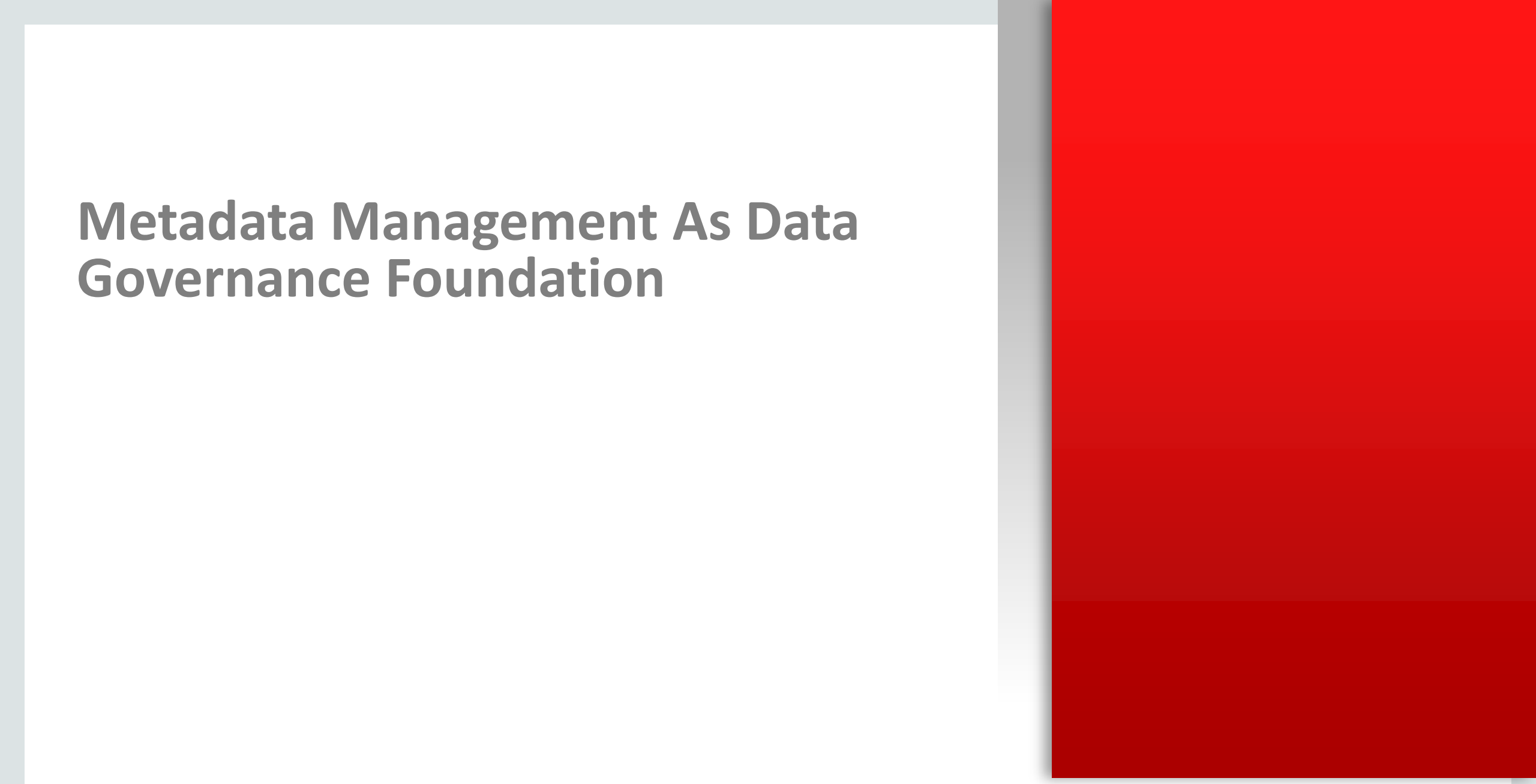
View Last 3 Months go



Switch to Count Mode
Show Passes on Graph
Show graph as percentages

Done

Metadata Management As Data Governance Foundation



Analytical And Operational Data Integration Solutions

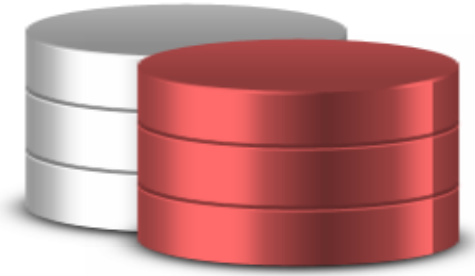
Oracle GoldenGate
Real Time Replica Between Any Databases

Active Data Guard
Disaster Recovery Of Oracle DB EE

Oracle Data Integrator
ETL Platform

Enterprise Data Quality
Data Quality Platform

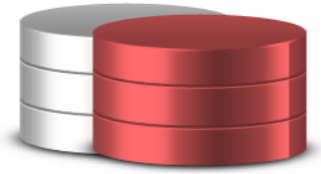
Enterprise Metadata Manager
Data Quality Platform



Migrations & Consolidations

OEMM
ODI
OEDQ
OGG

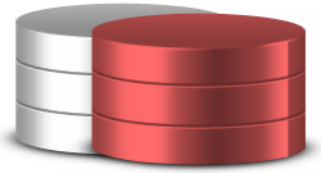
Zero Downtime
Migrations & Upgrades



New DB/HW/OS/APP

OGG

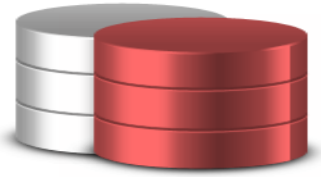
Active/Active High
Availability



Fully Active DB

OGG
ADG

Query Off-Loading
and Disaster Recovery

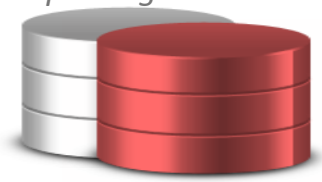


Reporting Database and/or DR database

OGG

OEMM
ODI
OEDQ

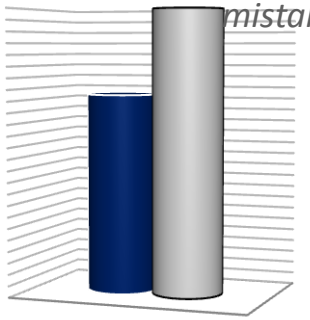
BI&DW
Synchronization and Loading



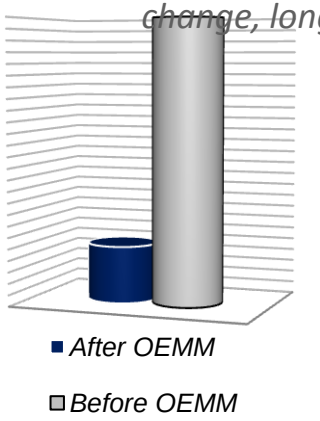
Data Warehouse

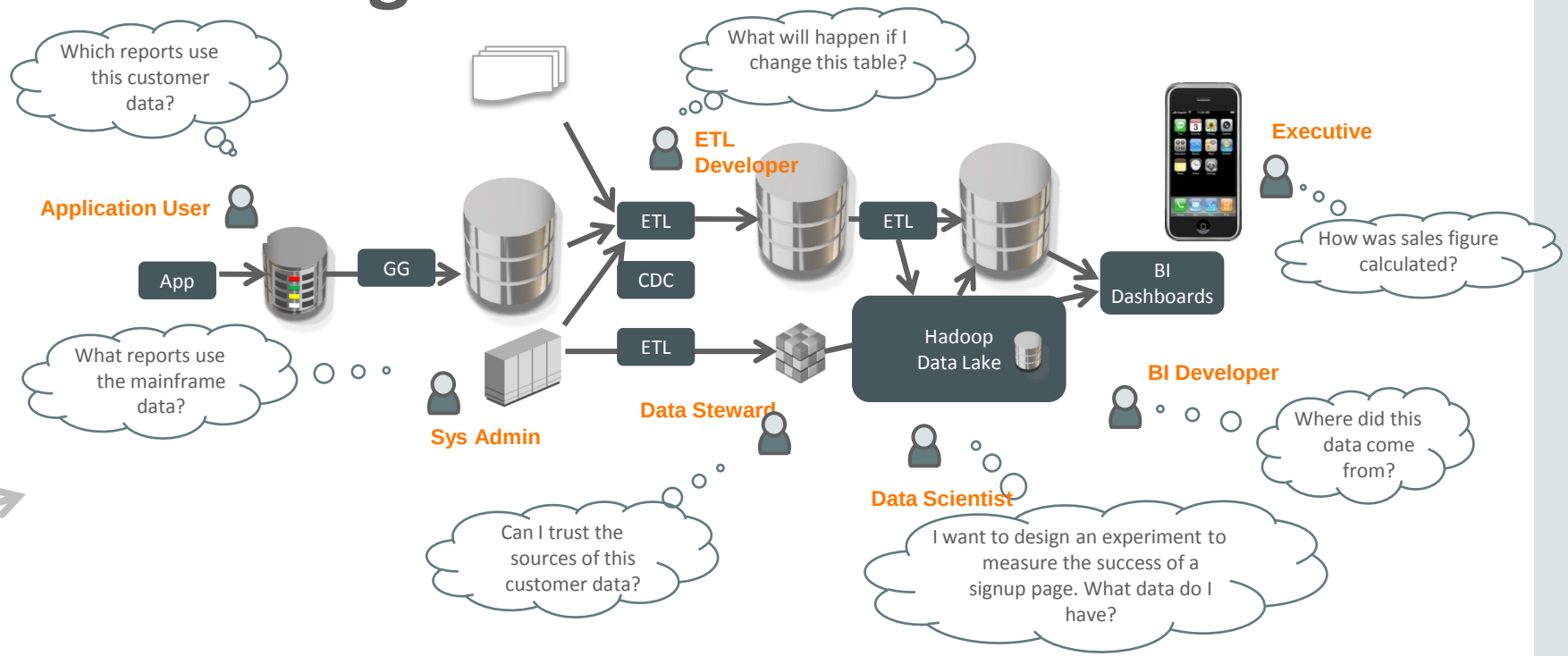
Value Of Metadata Management

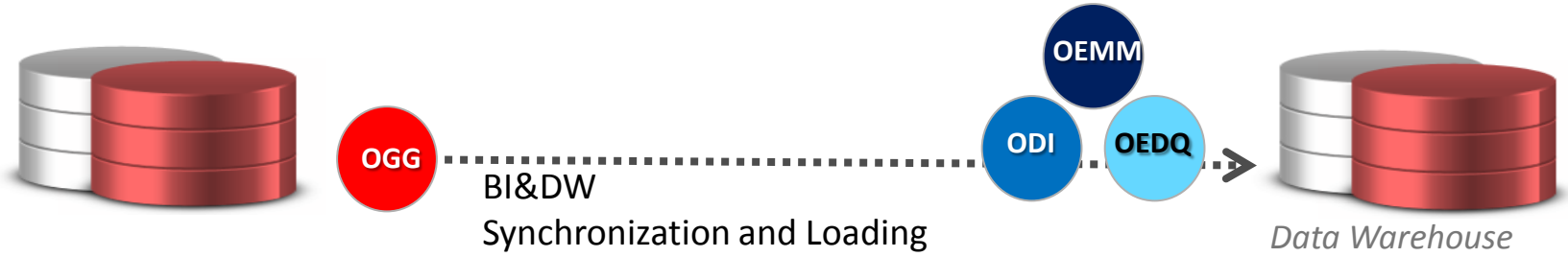
***CAPEX** - Reduce analytical project costs by 30% (not anymore ..post remediation costs, unnecessary mistakes)*



***OPEX** - Reduce data maintenance costs by 80% (not anymore.. hard to change, long to find)*







Metadata

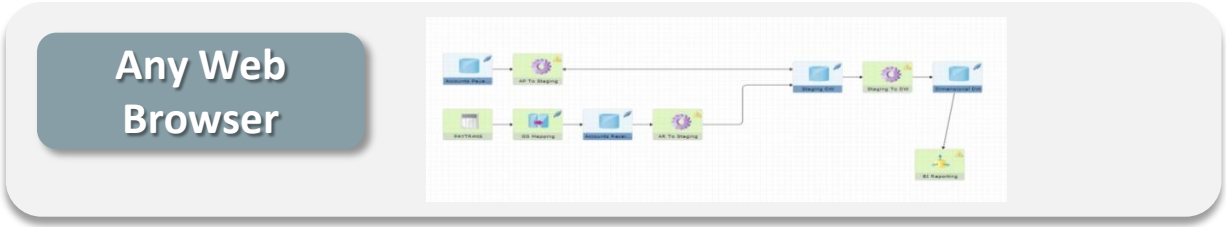
1. Title - the title of the data object
2. Creator - the person or entity responsible for creating the data object
3. Subject - subject terms or keywords that describe the data object
4. Description - a brief description, or abstract, of the data object
5. Publisher - the entity responsible for making the data object available
6. Contributor - a person or entity who contributed to the creation of the data object
7. Date - data of creation, publication, or revision of the data object
8. Type - the type of object. For data this would typically be "dataset"
9. Format - a description of the format or file type(s) of the data object
10. Identifier - a permanent identifier used to locate and identify the data object
11. Language - the language(s) used within the data object (if applicable)
12. Source - a relational element describing the lineage of the data object
13. Relation - a relational element describing the relationship of this data object to other objects, collections, or entities
14. Coverage - describes the spatial and temporal context of the data object
15. Rights - describes any rights, restrictions, or terms of use

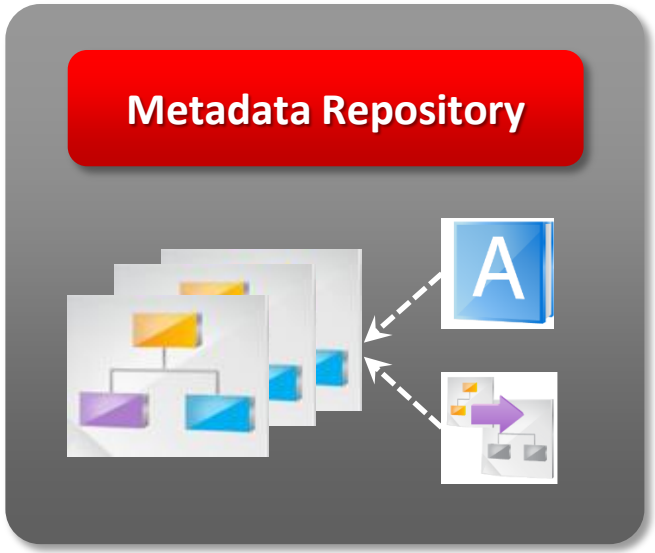
Entity Name: Customer
Alias Names: Account, Client

Definition:	A person or an organization that purchases goods or services from the company.
Remarks:	Customer entity includes regular, current, and past customers.
Source Systems:	Finished Goods Orders, Maintenance Contracts, Online Sales.
Create Date:	January 15, 1999
Last Update Date:	January 21, 2001
Update Cycle:	Weekly
Last Full Refresh Date:	December 29, 2000
Full Refresh Cycle:	Every six months
Data Quality Reviewed:	January 25, 2001
Last Deduplication:	January 10, 2001
Planned Archival:	Every six months
Responsible User:	Jane Brown

Figure 9-1 Metadata element for *Customer* entity.

OEMM Architecture





Analytical And Operational Data Integration Solutions

Operational Integration (Movement / Transformation)

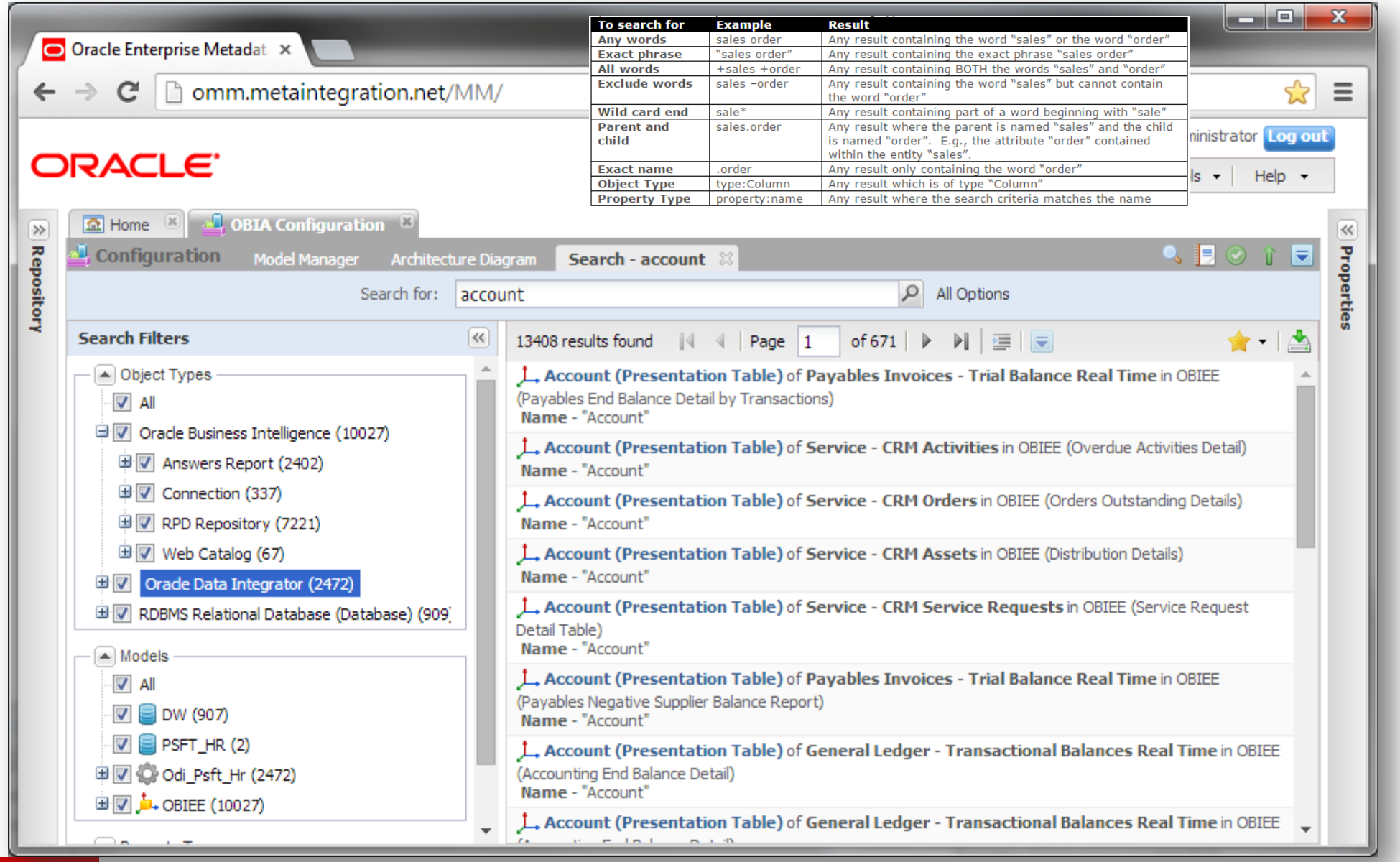
- Hadoop HBase
- Hadoop Hive/Flume
- HP Enscribe
- HP NonStop
- HP Neoview
- Hypersonic SQL
- IBM DB2 i Series
- IBM DB2 UDB
- IBM DB2 z Series
- IBM Informix
- IBM Netezza
- JMS / MQ
- Microsoft Access
- Microsoft SQLServer
- MySQL
- Pivotal Greenplum
- PostgreSQL
- Salesforce.com
- SAP BW / BI
- SAP ERP / ECC
- SAS
- SQL/MP
- SQL/MX
- Sybase ASE
- Sybase IQ
- Teradata
- Oracle Database
- Oracle Exadata
- Oracle Big Data Appliance
- Oracle TimesTen
- Oracle OLAP
- Oracle Business Intelligence
- Oracle BI Applications
- Oracle E-Business Suite
- Oracle JD Edwards Enterprise One
- Oracle JD Edwards World
- Oracle Fusion Applications
- Oracle Governance Risk and Compliance
- Oracle Fusion AIA
- Oracle Retail Applications
- Oracle Agile BI / DW
- Oracle Agile PLM for Process
- Oracle iFlex FlexCUBE
- Oracle iFlex Mantas
- Oracle Hyperion Applications
- Oracle PeopleSoft
- Oracle Siebel CRM / OnDemand
- Oracle Communications
- Oracle WebLogic Server
- Oracle Coherence Data Grid
- Oracle SOA Suite
- Oracle Enterprise Service Bus

Metadata Harvesting (Glossary, Lineage & Impact Analysis)

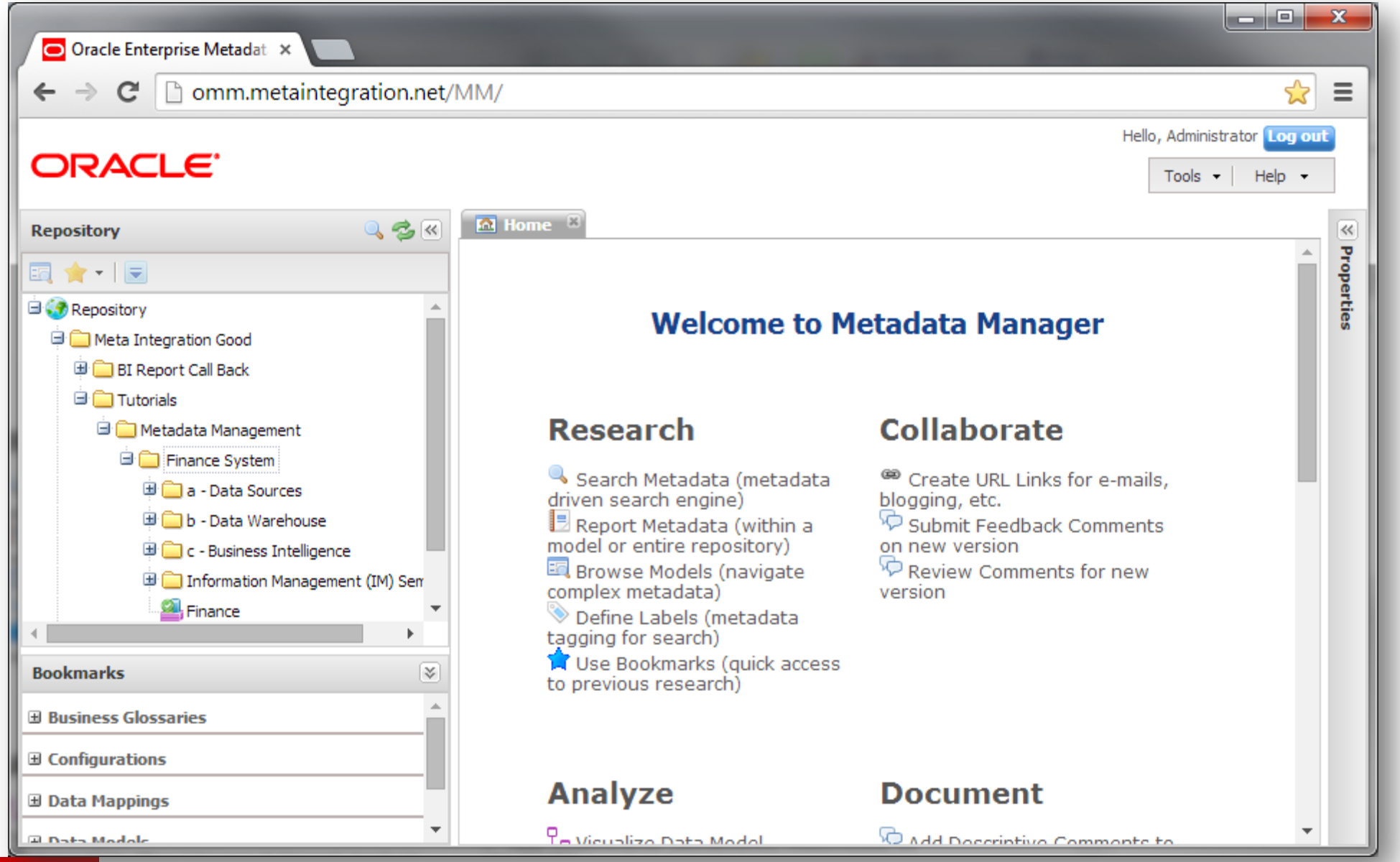
- Adaptive
- Altova
- Apache Hcatalog
- Apache Hive/HQL
- Borland
- CA ERwin
- Cloudera Impala
- COBOL Copybook
- DataStax
- Embarcadero
- EMC ProActivity
- GentleWare
- Google BigQuery
- Grandite
- Hadapt Hive
- Hortonworks Hive
- IBM Cognos
- IBM DB2
- IBM DataStage
- IBM Discovery
- IBM Federation Server
- IBM Lotus Notes
- IBM Netezza
- IBM Rational Rose
- IBM Rational Architect
- Informatica Metadata Mgr.
- Informatica PowerCenter
- CoSORT
- ISO SQL Standard (DDL)
- MapR Hadoop Hive
- MicroFocus
- Microsoft Access
- Microsoft Office Excel
- Microsoft Visio
- Microsoft SQL Server
- Microsoft SSIS
- Microsoft Visual Studio
- Microstrategy
- Magic Draw
- OMG CWM Standard
- OMG UML Standard
- Oracle BI Answers
- Oracle BI Enterprise Edition
- Oracle BI Server
- Oracle DAC
- Oracle Data Integrator
- Oracle Data Modeler
- Oracle Database
- Oracle Designer
- Oracle Hyperion Applications
- Oracle Hyperion Essbase
- Oracle Warehouse Builder
- Pivotal Greenplum
- PostgreSQL
- QlikView
- SAP BO Crystal Reports
- SAP BO Designer
- SAP BO Desktop Intelligence
- SAP BO Repository
- SAP BO Data Integrator
- SAP BO Data Steward
- SAP Master Data Management
- SAP Sybase PowerDesigner
- SAP Sybase ASE Database
- SAS Data Integration Studio
- SAS BI Server
- SAS Information Map
- SAS Metadata Management
- SAS OLAP Server
- Select
- Sparx Architect
- Syncsort
- Tableau
- Talend
- Teradata
- Tigris
- Visible
- W3C DTD & XSD Schema

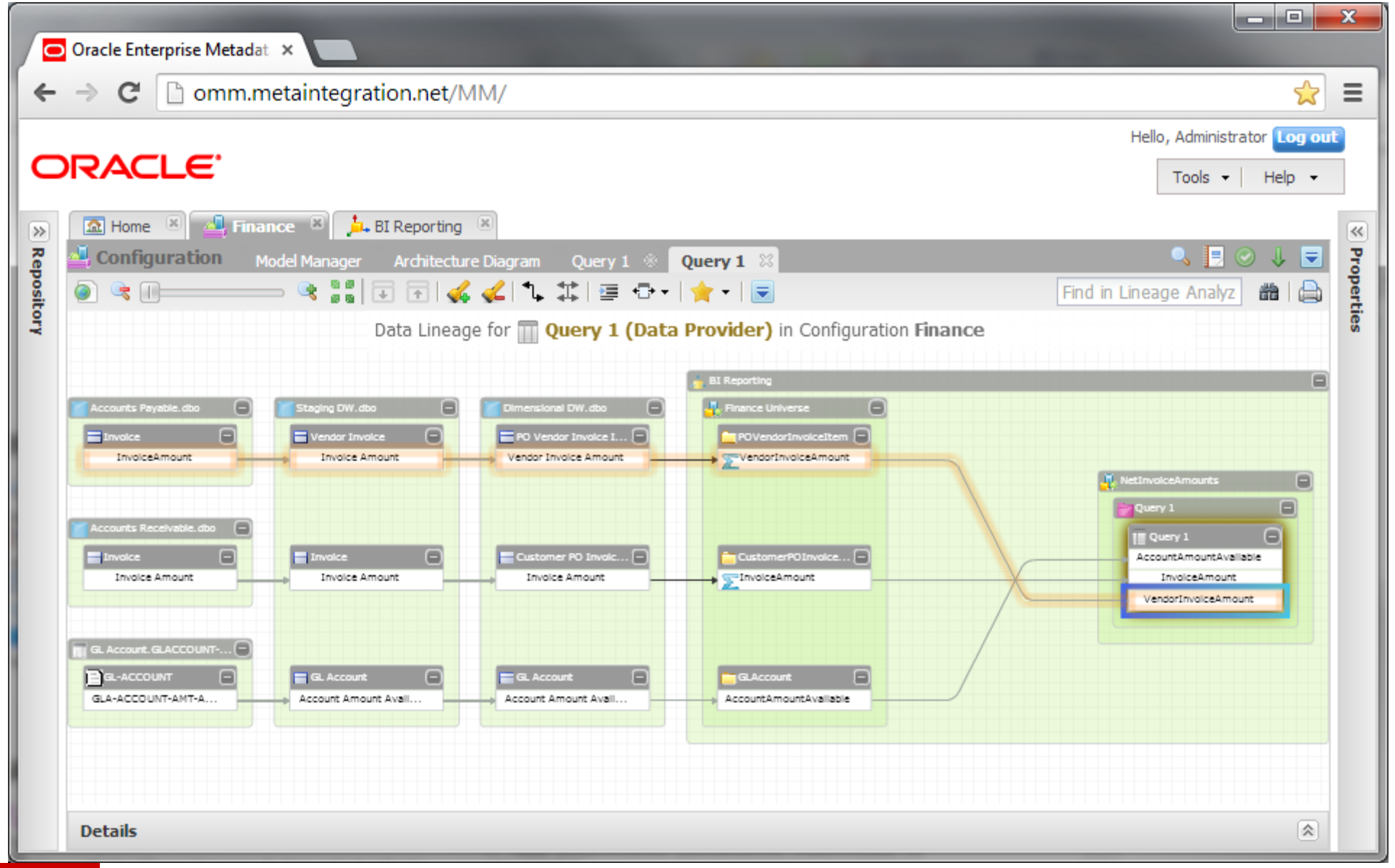
+ open APIs and standards based meta-model

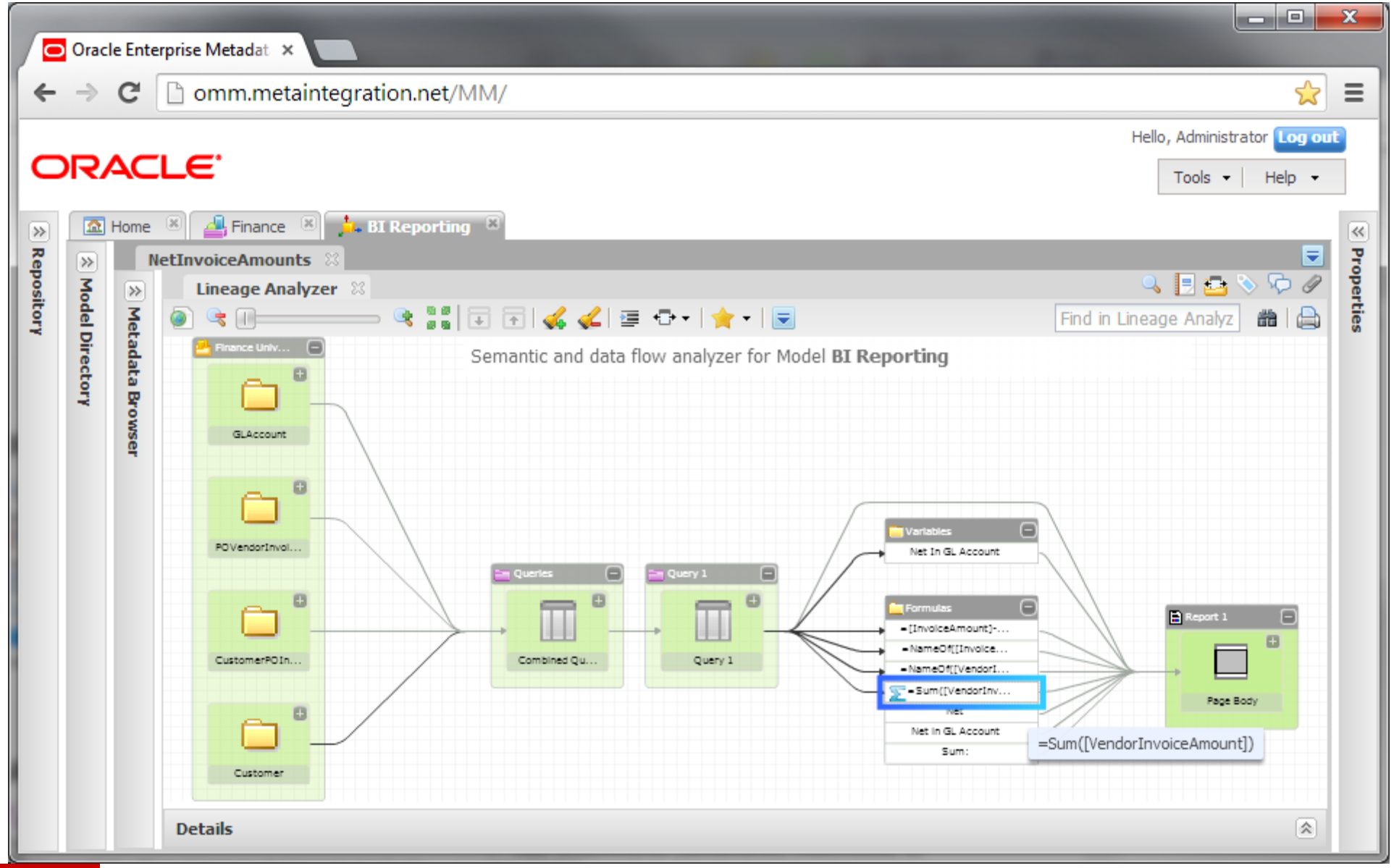
ORACLE

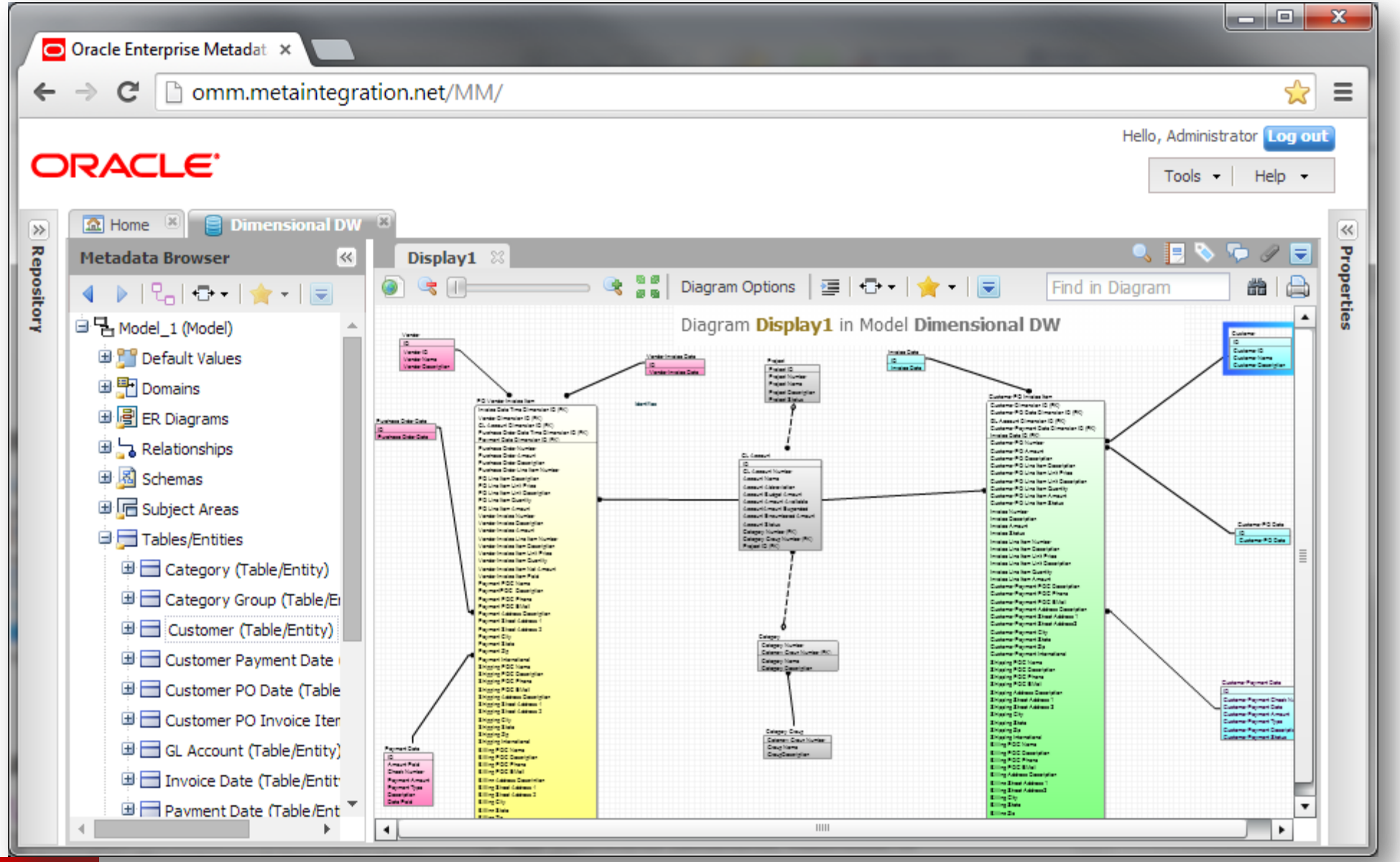


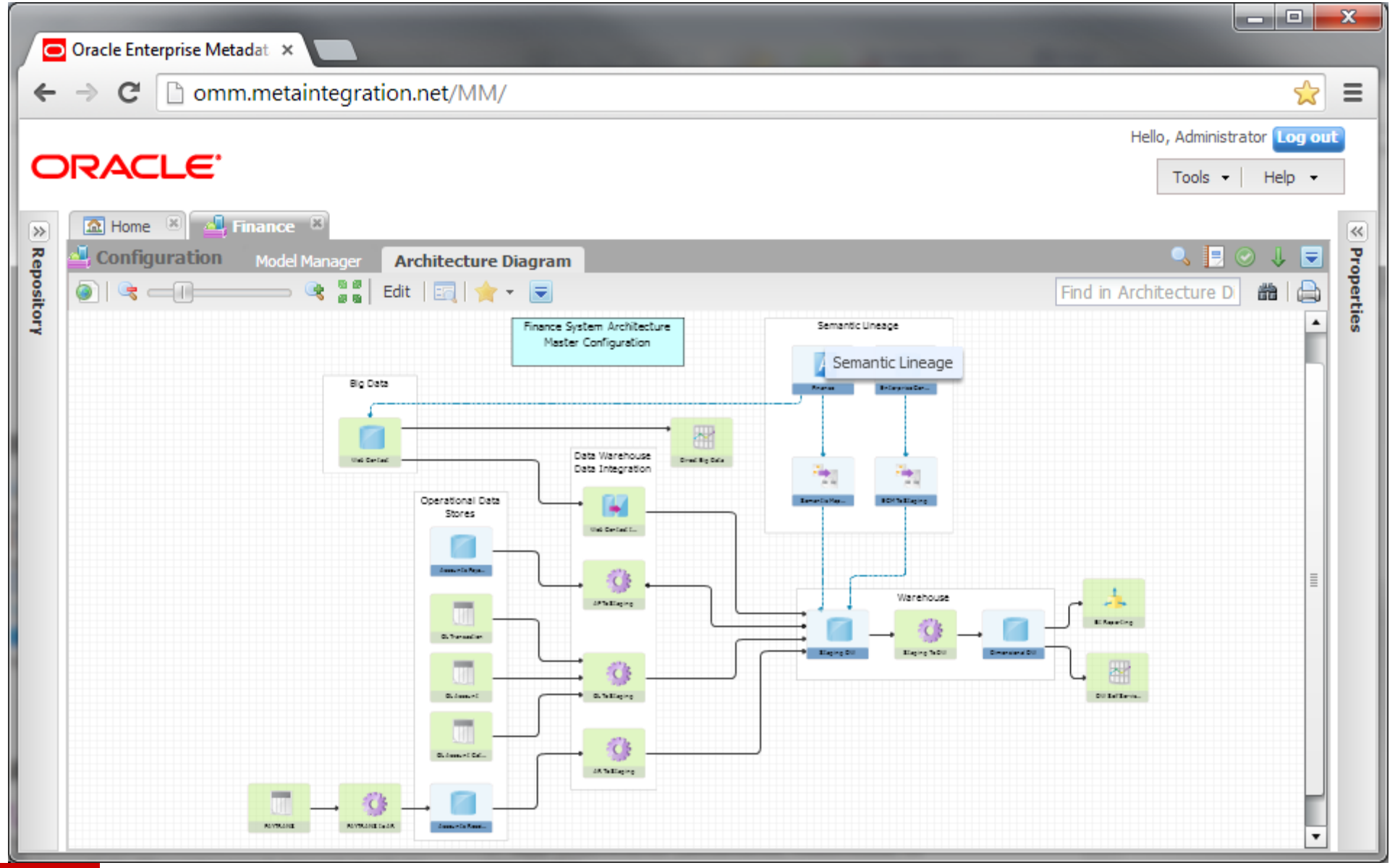
To search for	Example	Result
Any words	sales order	Any result containing the word "sales" or the word "order"
Exact phrase	"sales order"	Any result containing the exact phrase "sales order"
All words	+sales +order	Any result containing BOTH the words "sales" and "order"
Exclude words	sales –order	Any result containing the word "sales" but cannot contain the word "order"
Wild card end	sale*	Any result containing part of a word beginning with "sale"
Parent and child	sales.order	Any result where the parent is named "sales" and the child is named "order". E.g., the attribute "order" contained within the entity "sales".
Exact name	.order	Any result only containing the word "order"
Object Type	type:Column	Any result which is of type "Column"
Property Type	property:name	Any result where the search criteria matches the name

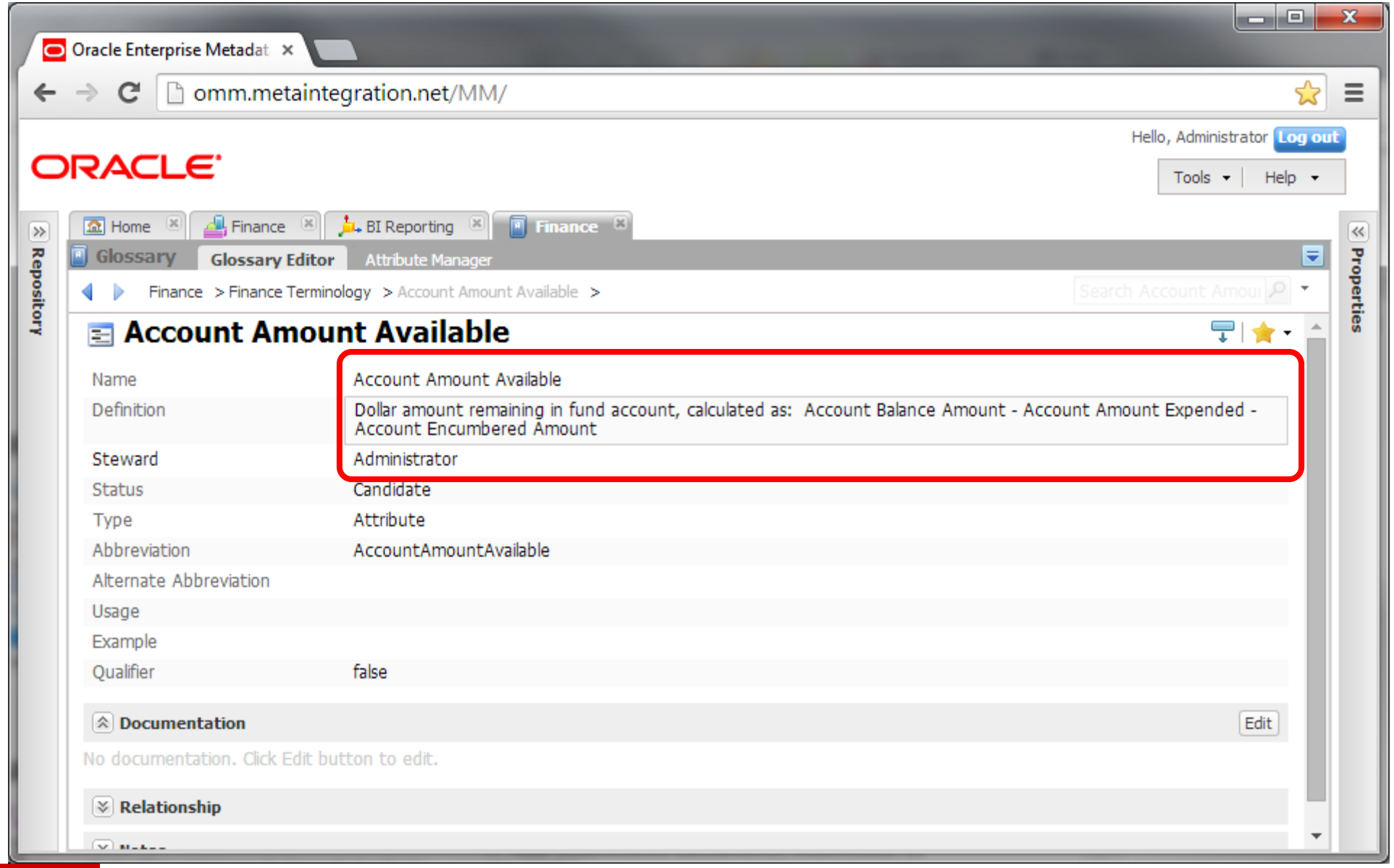


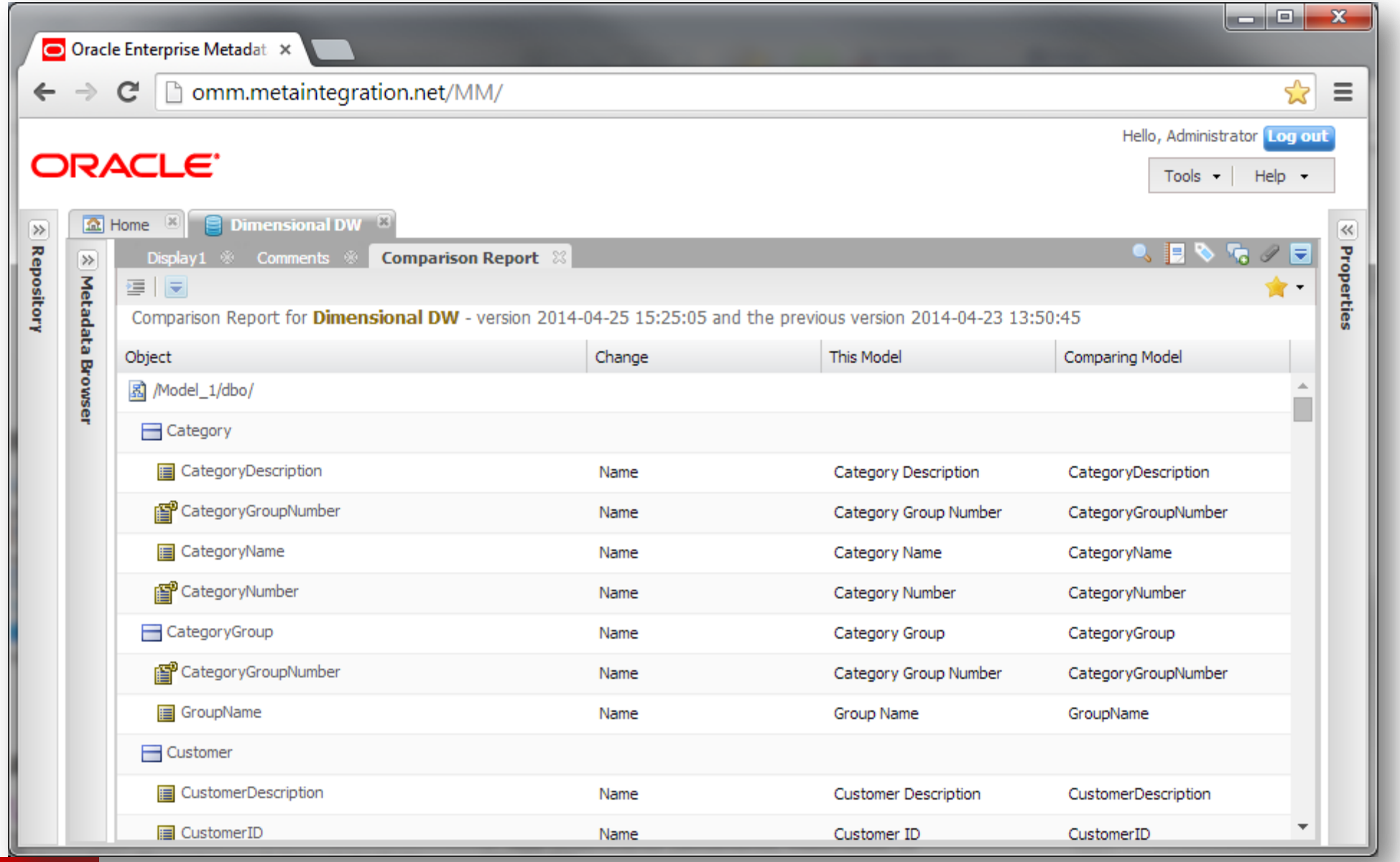


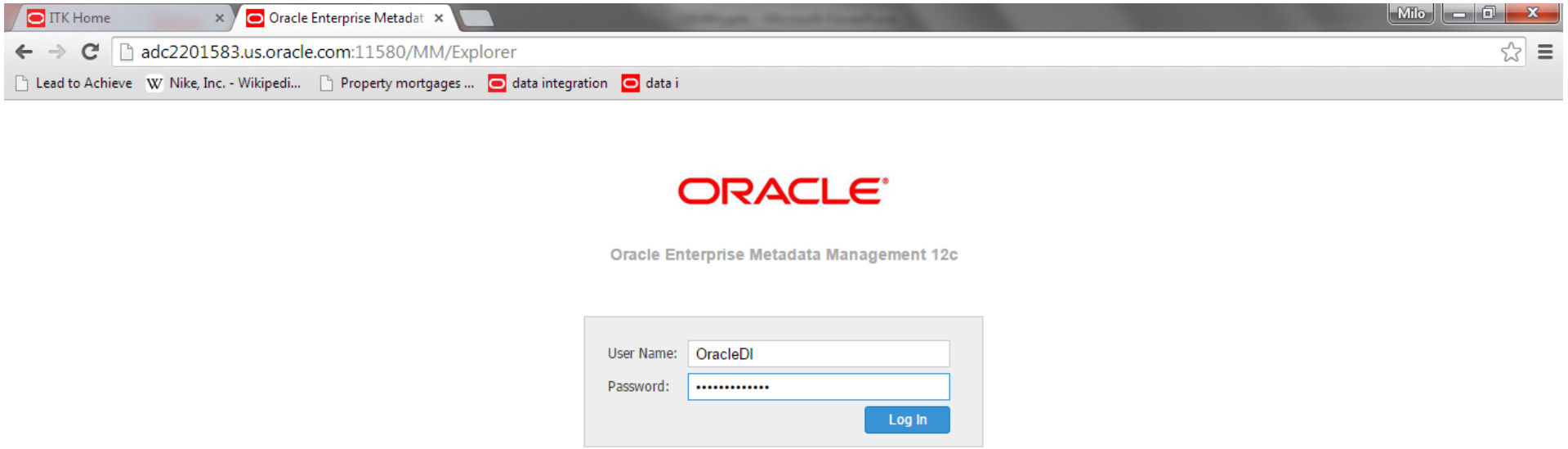


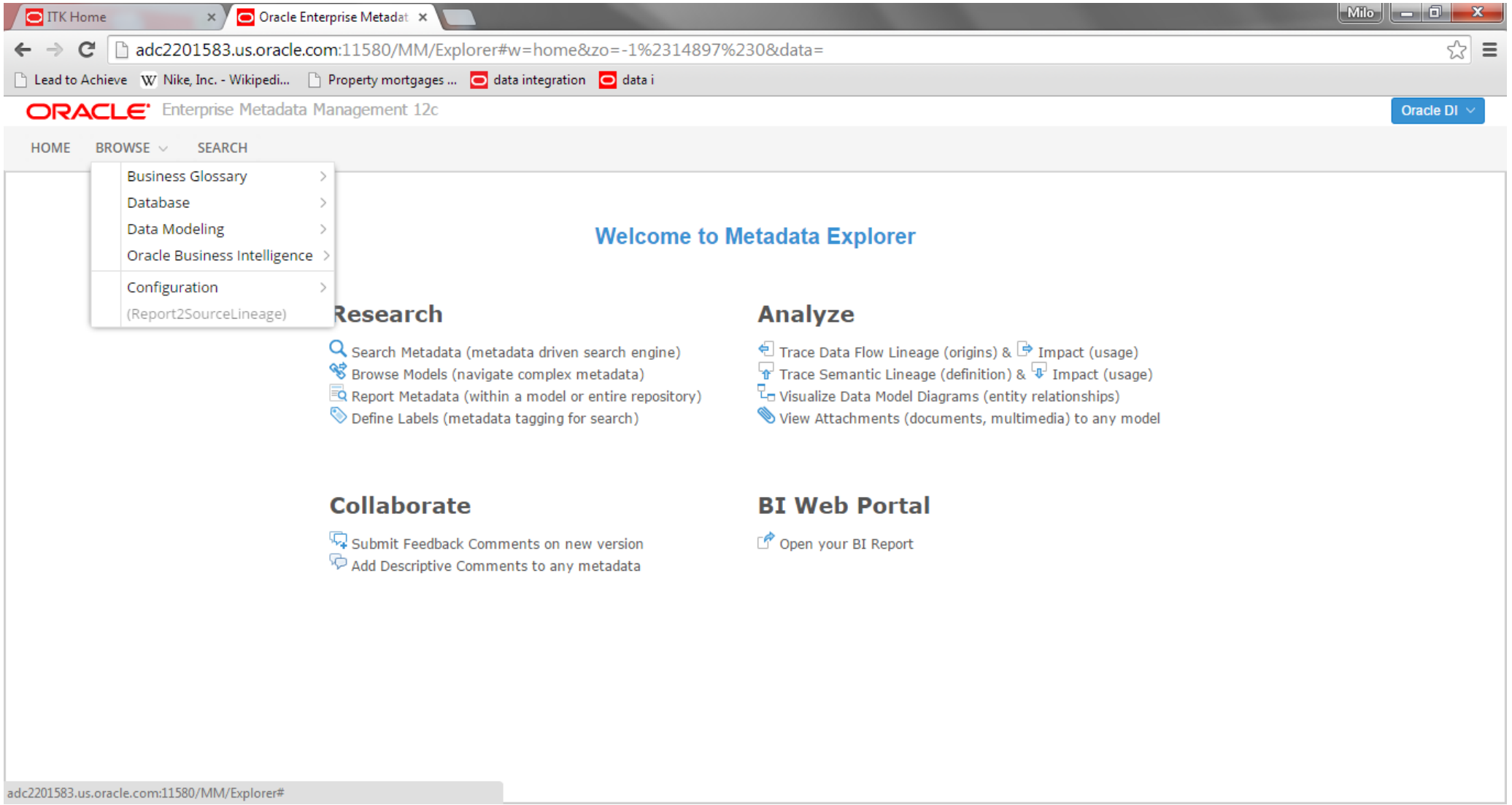


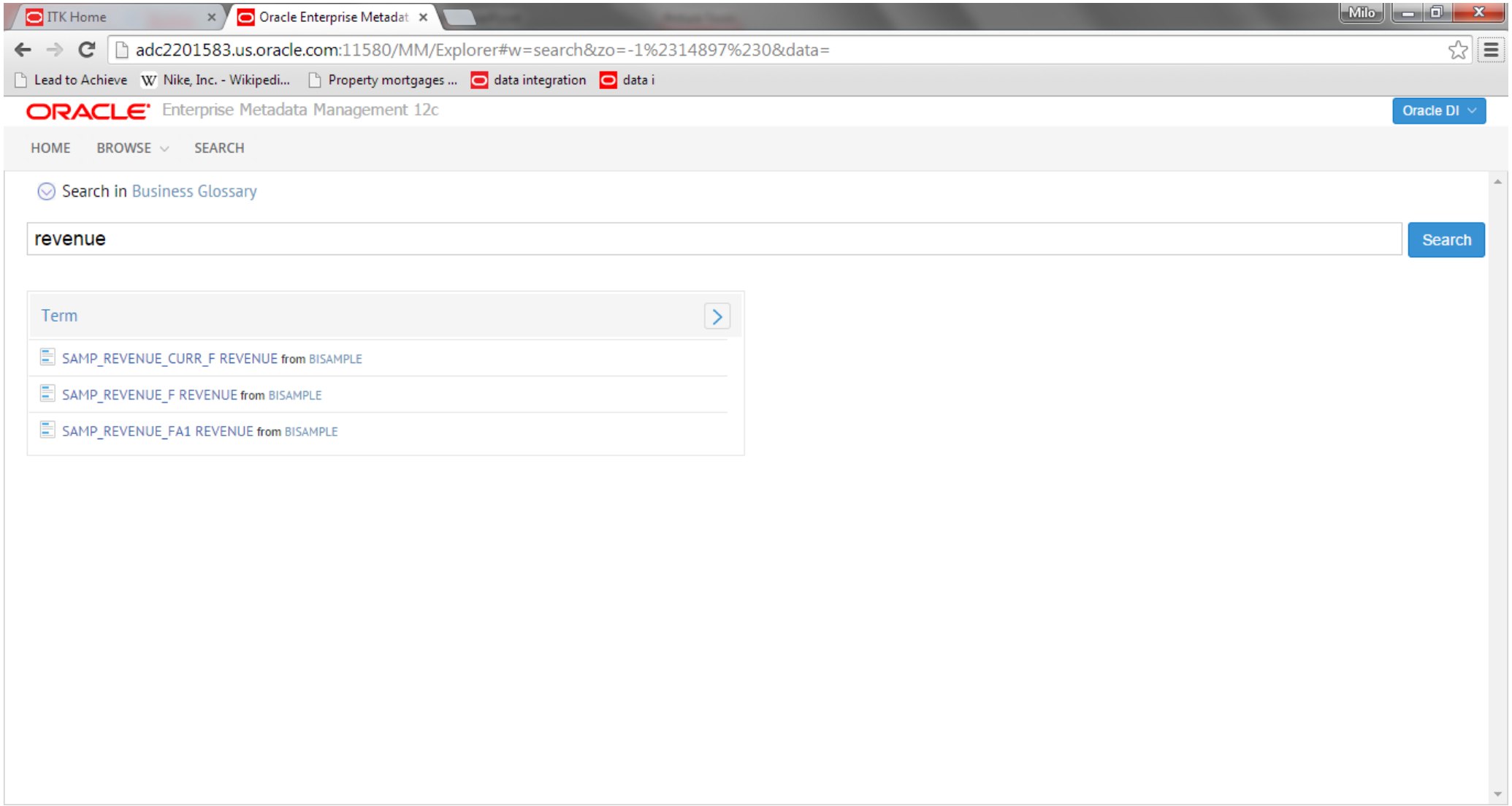


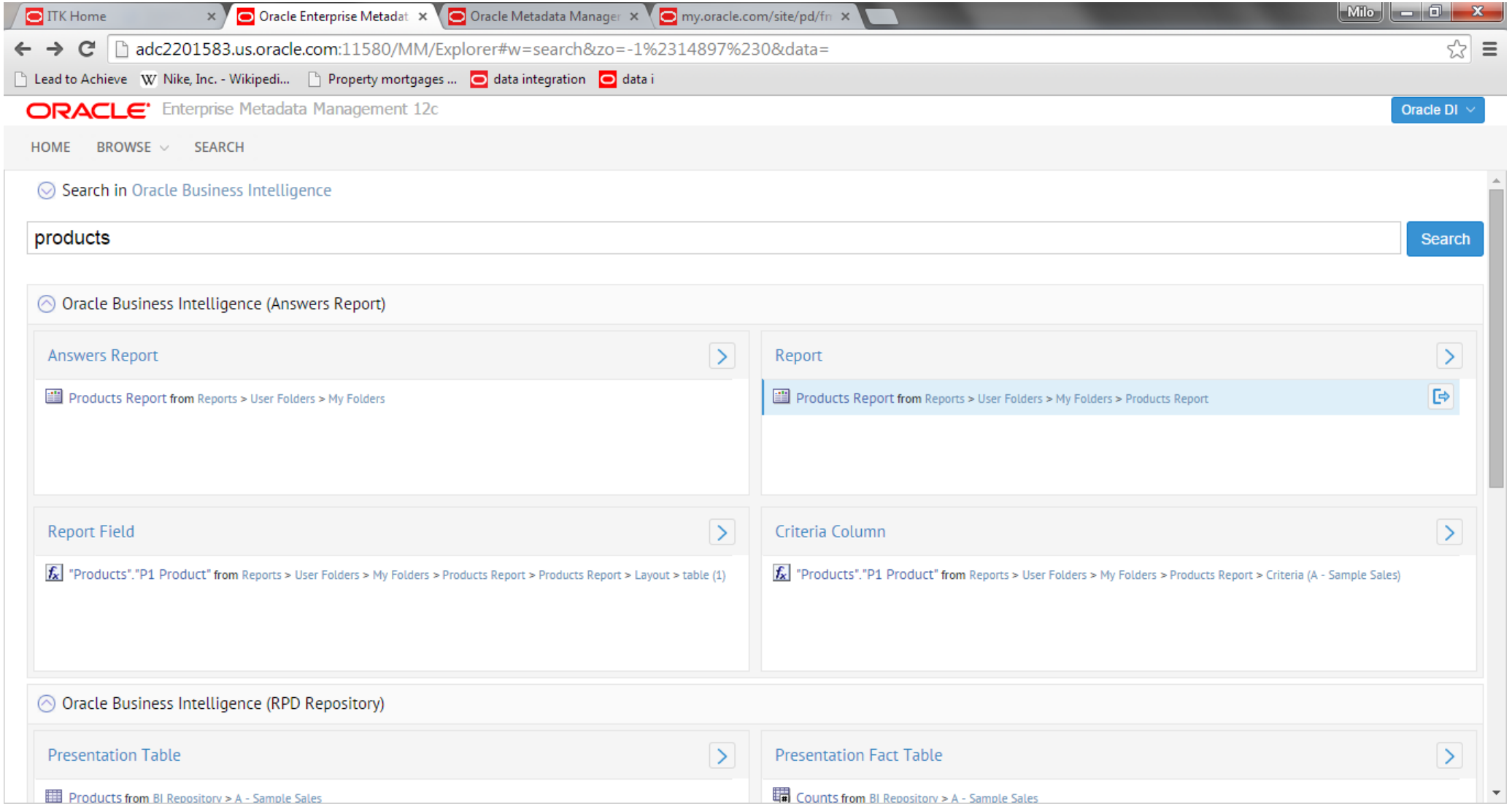


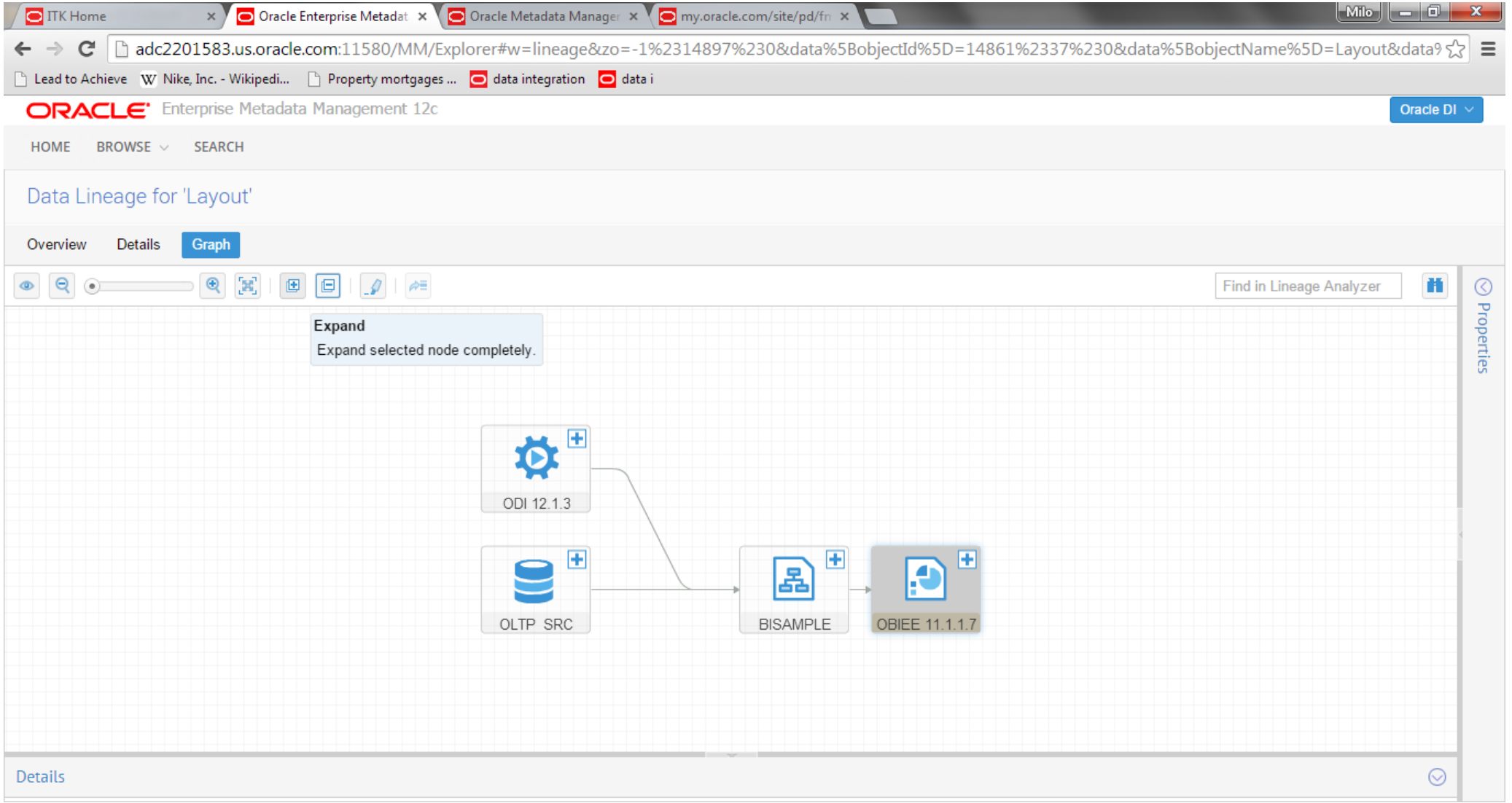


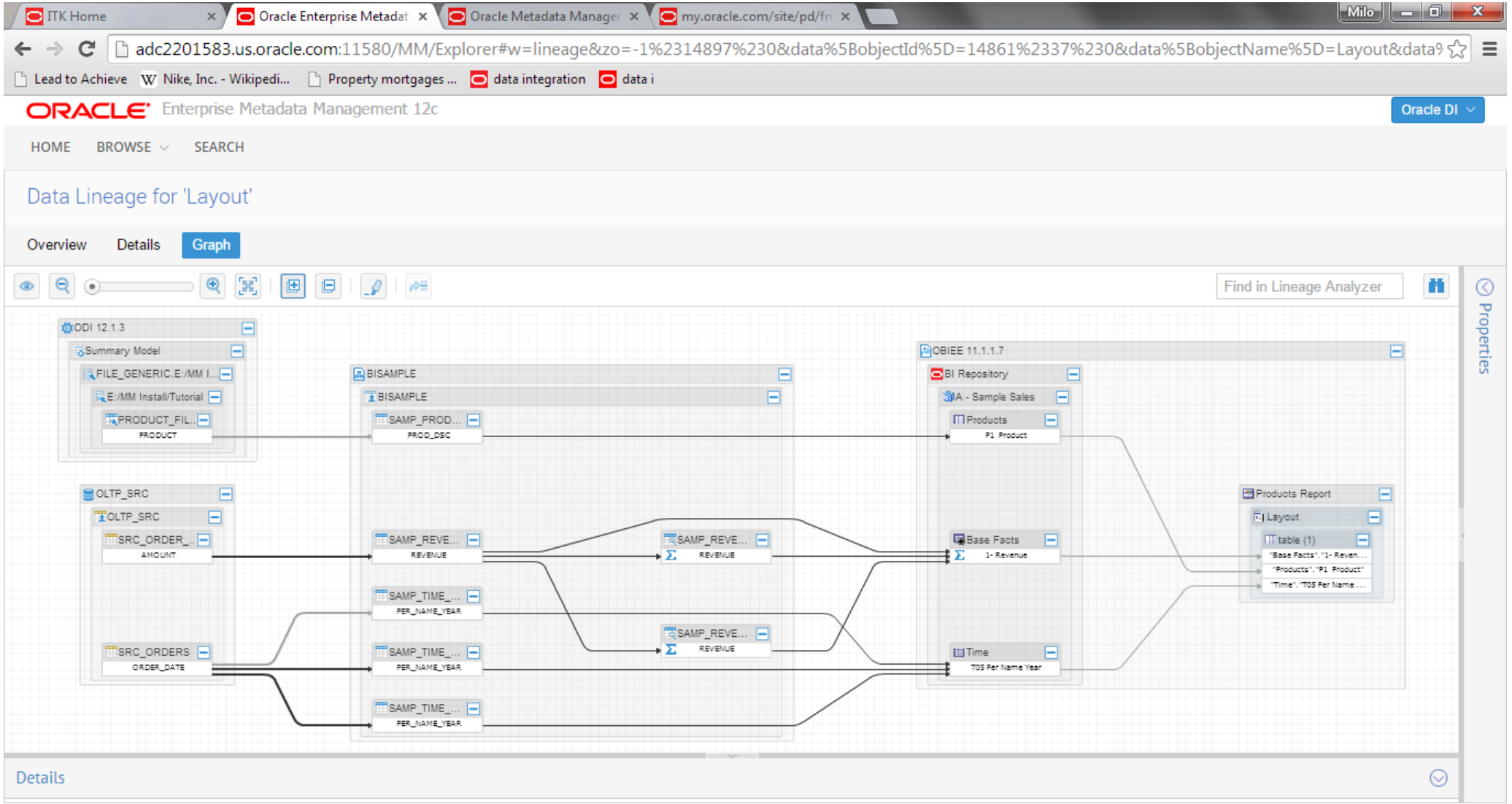


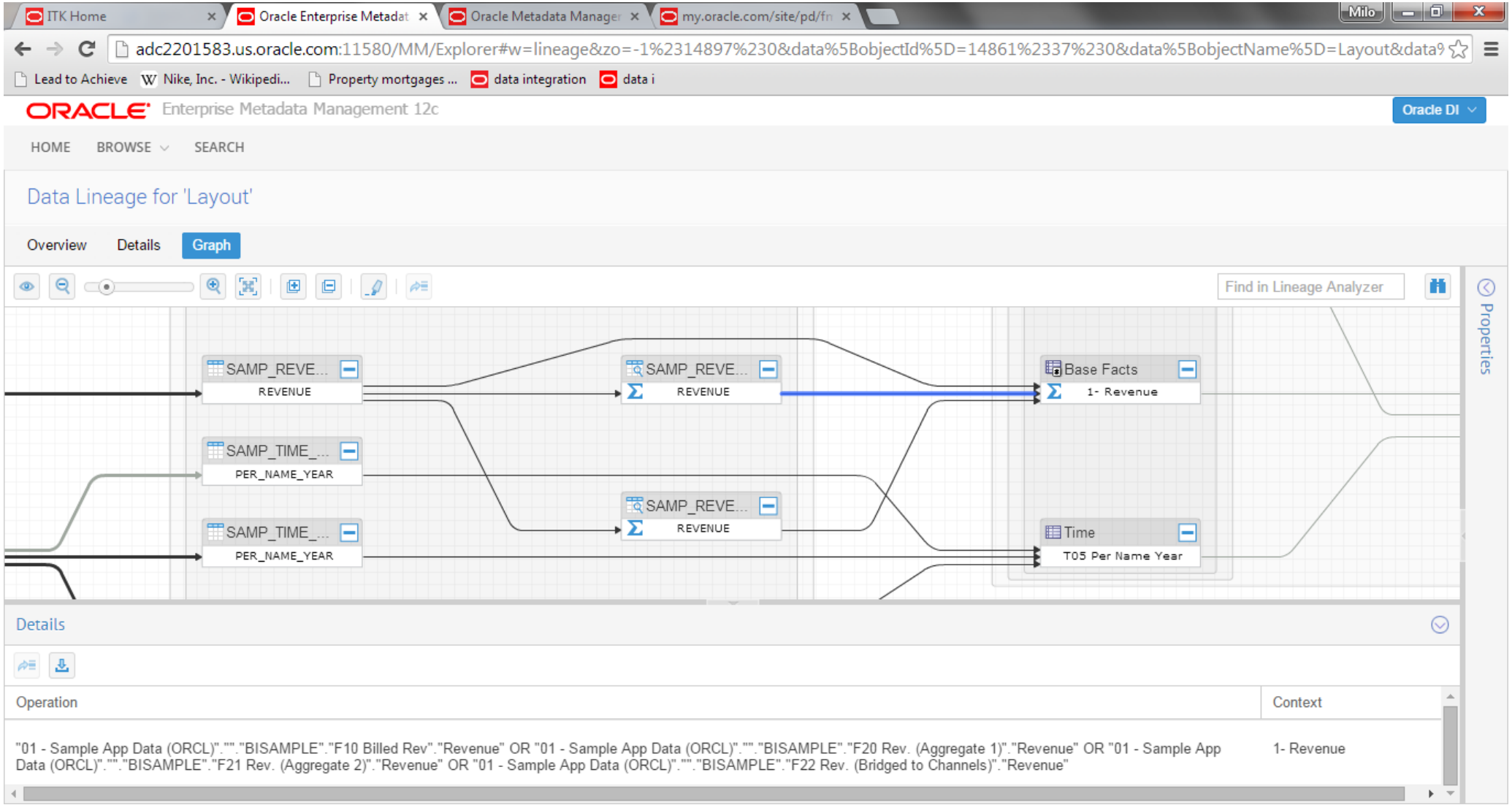


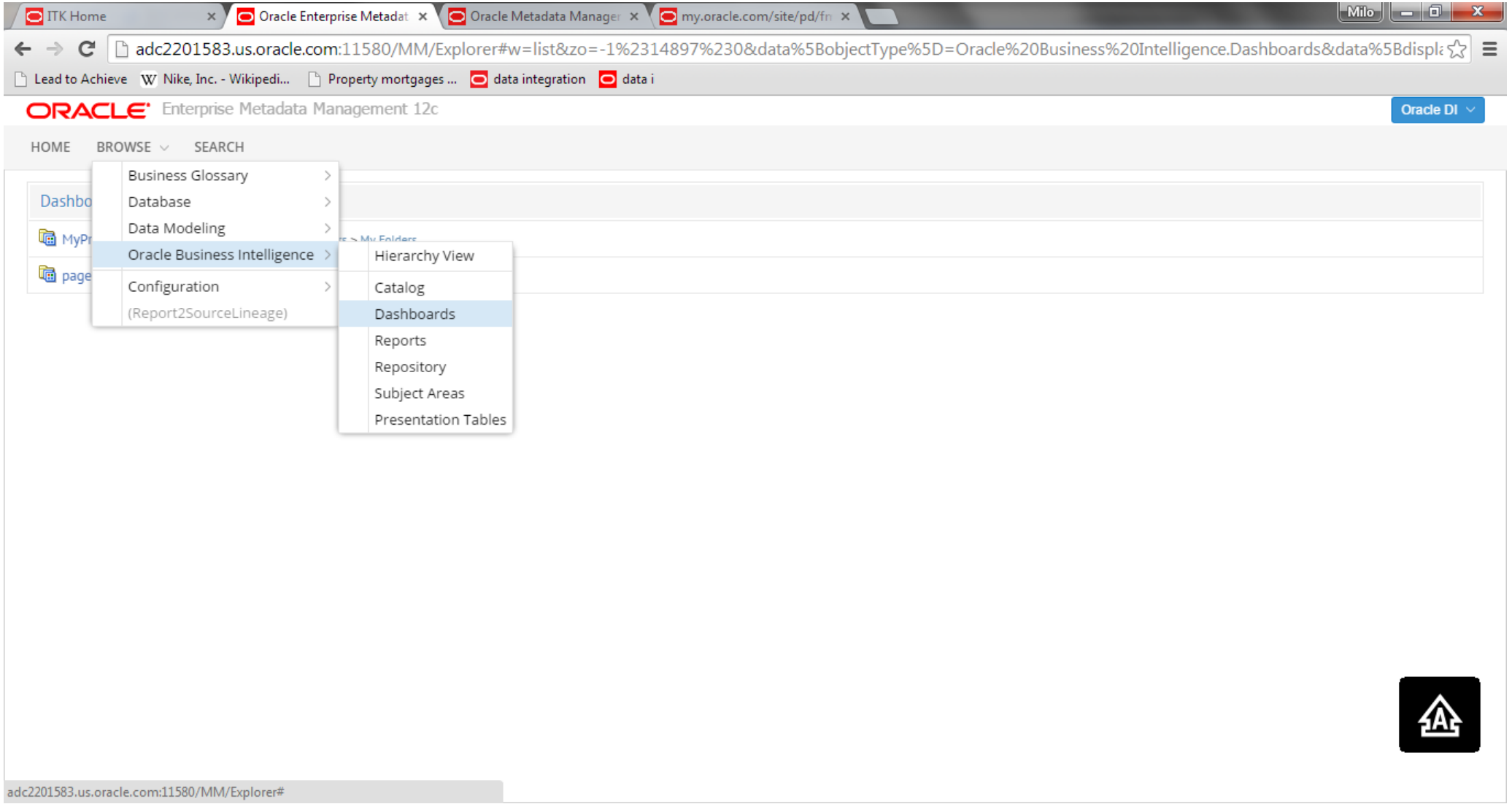


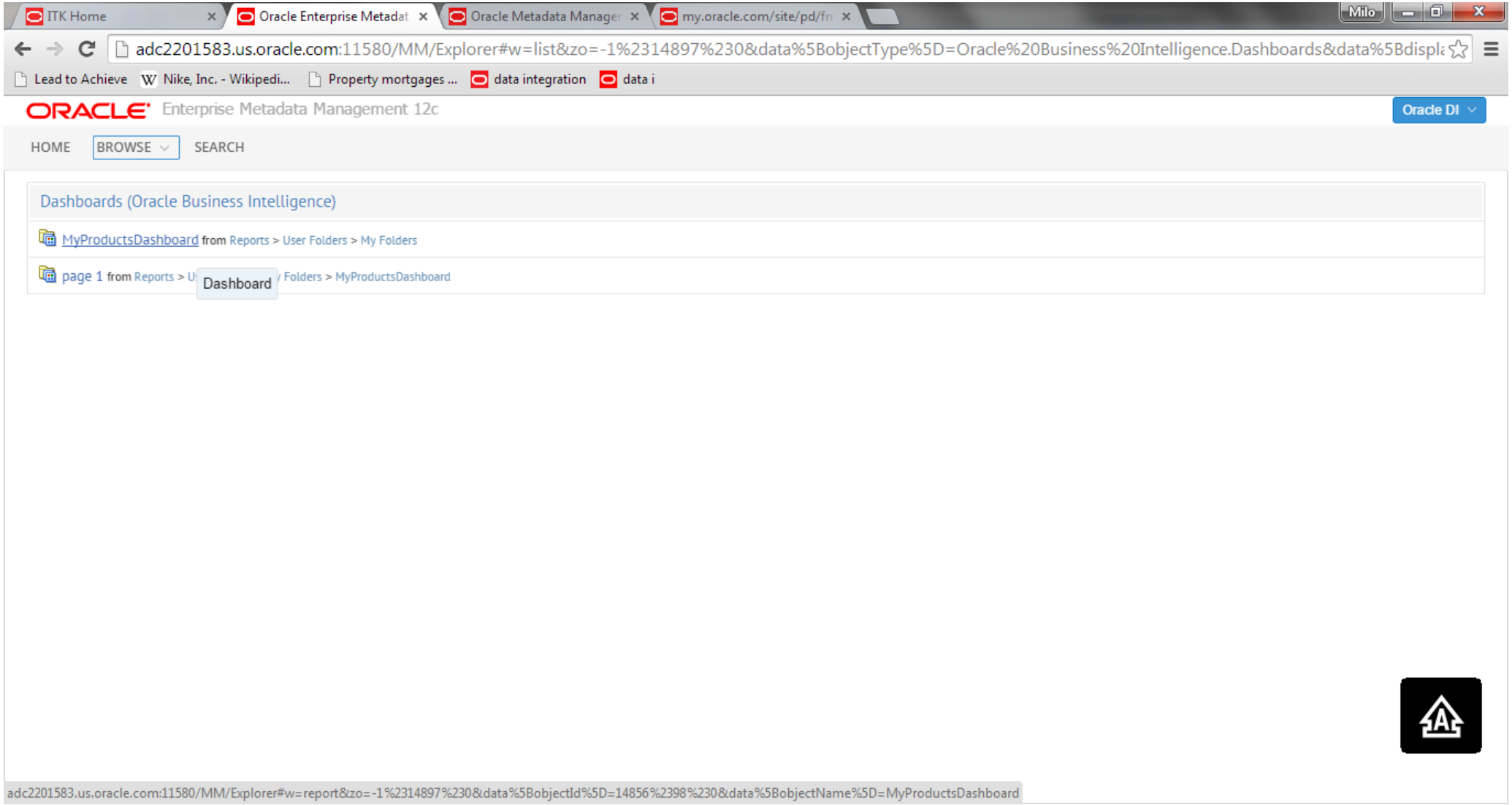


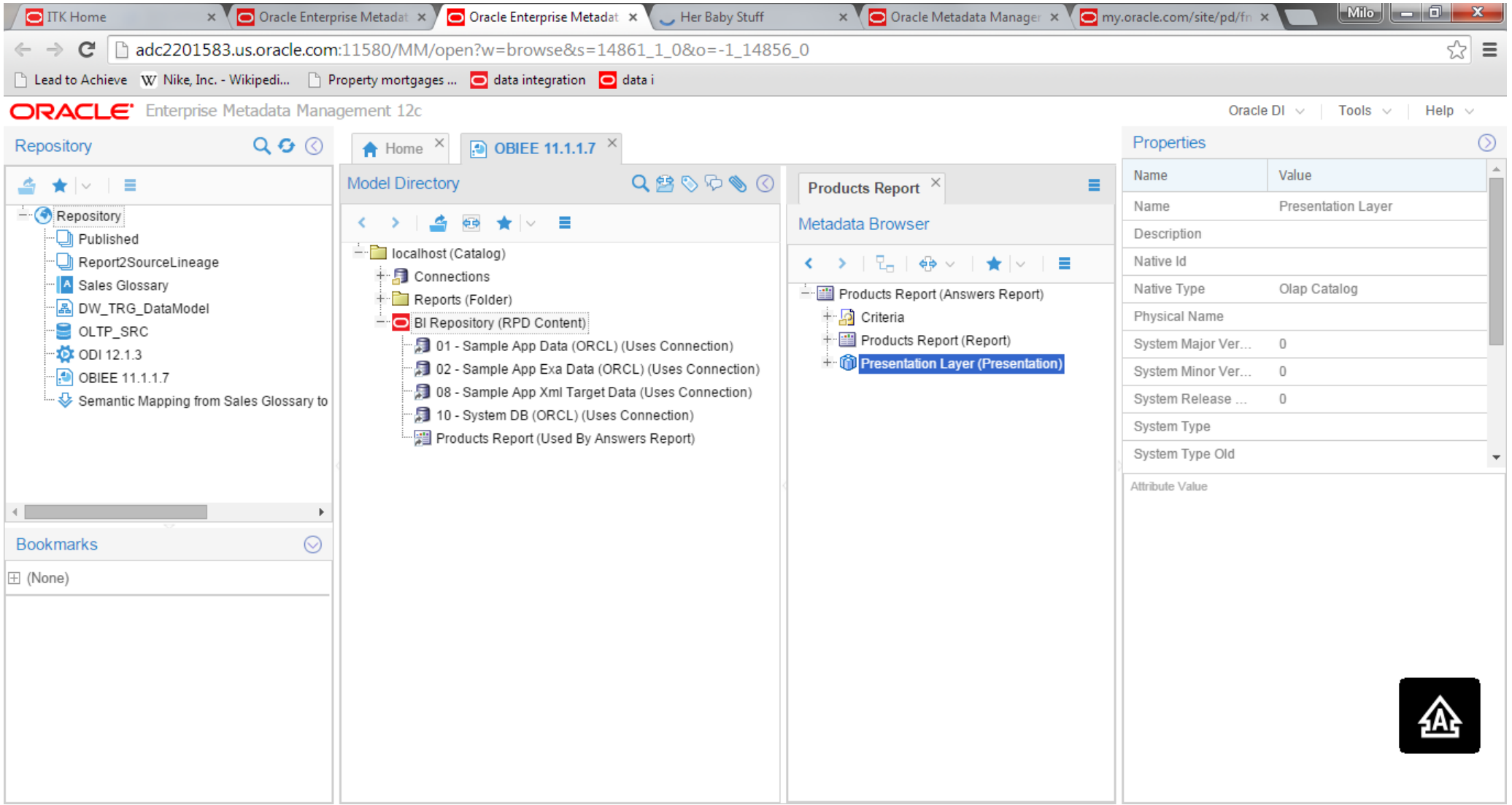


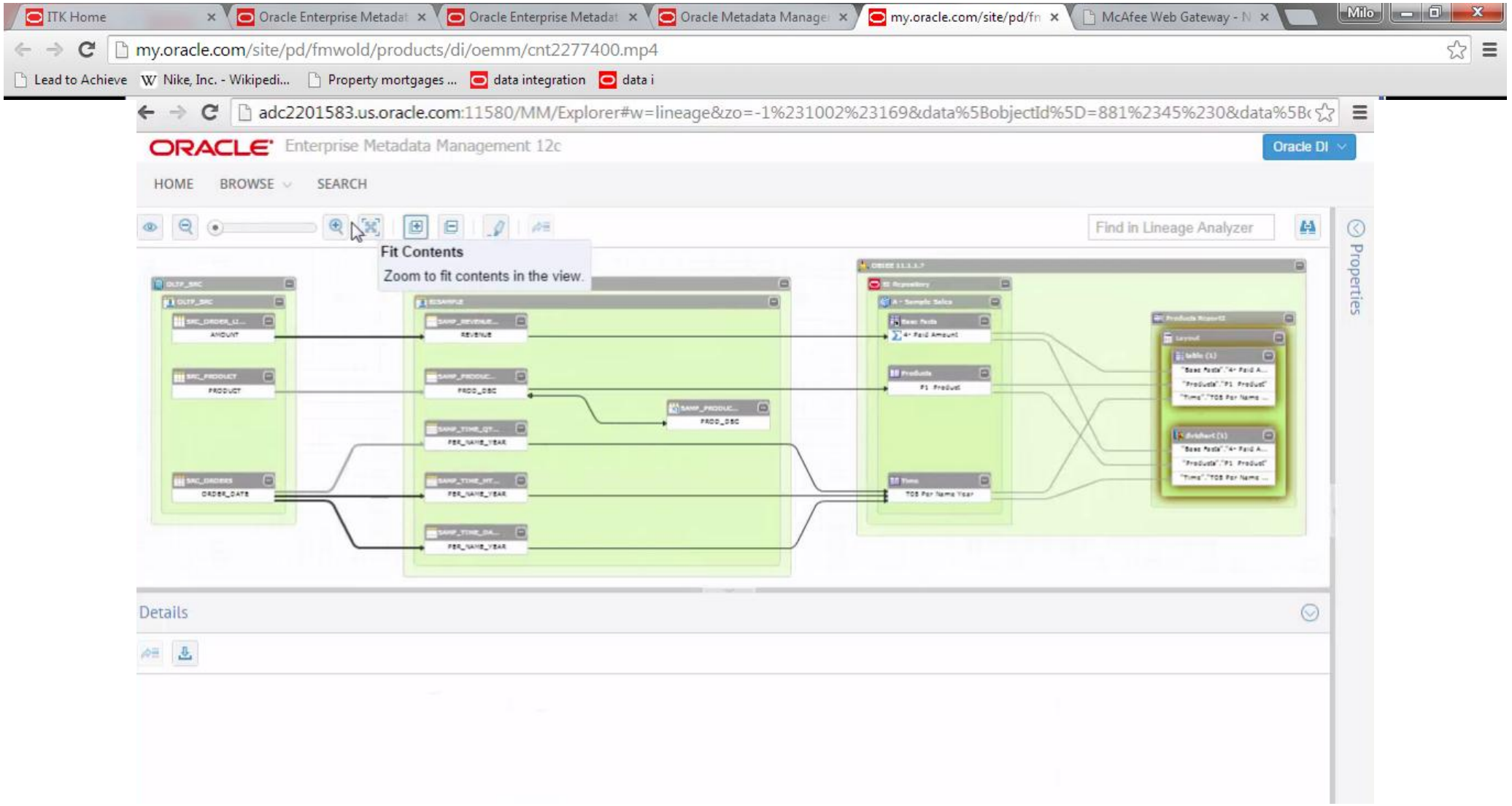


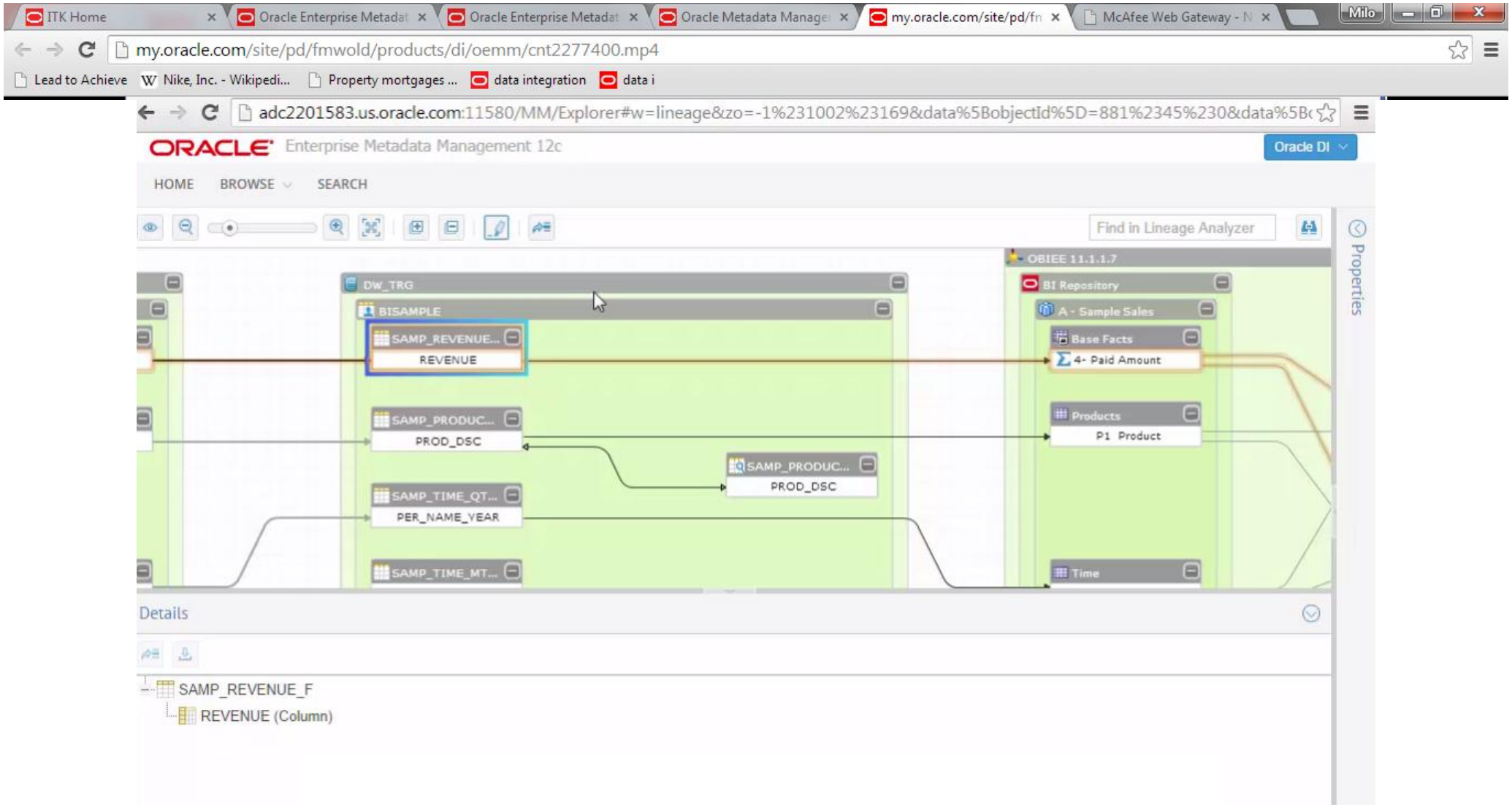


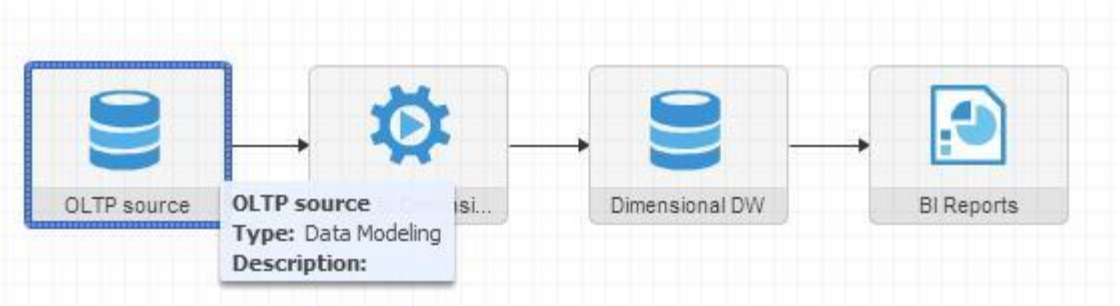


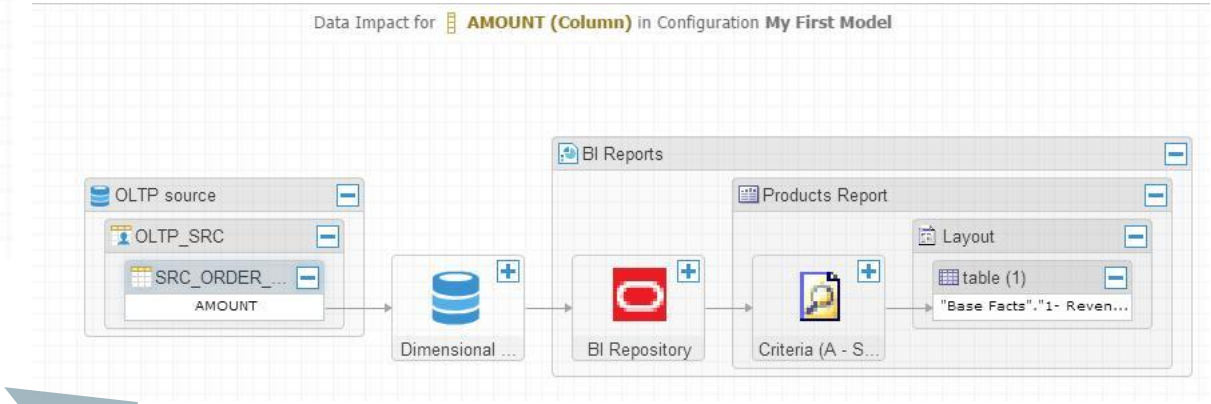


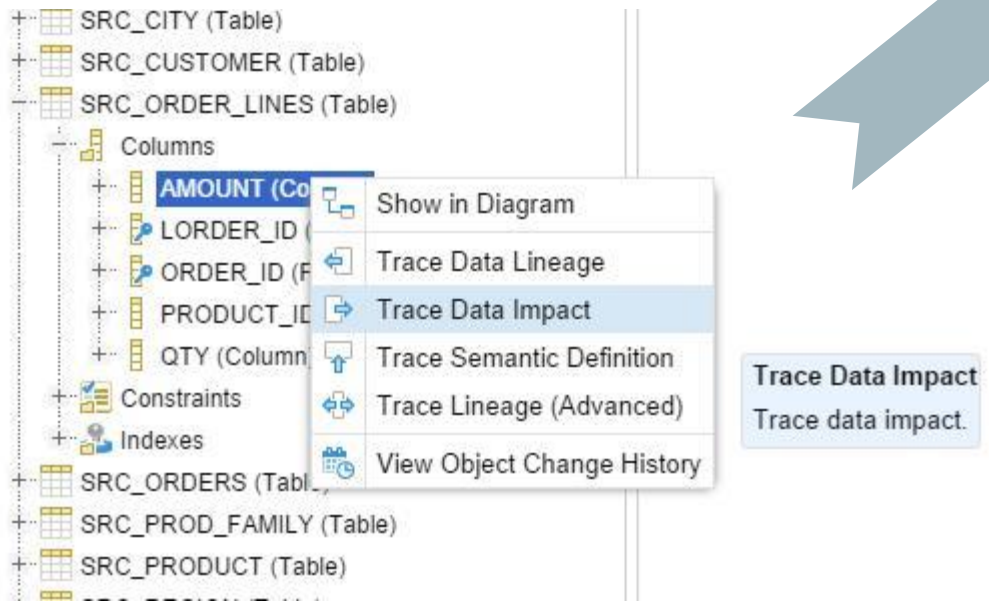




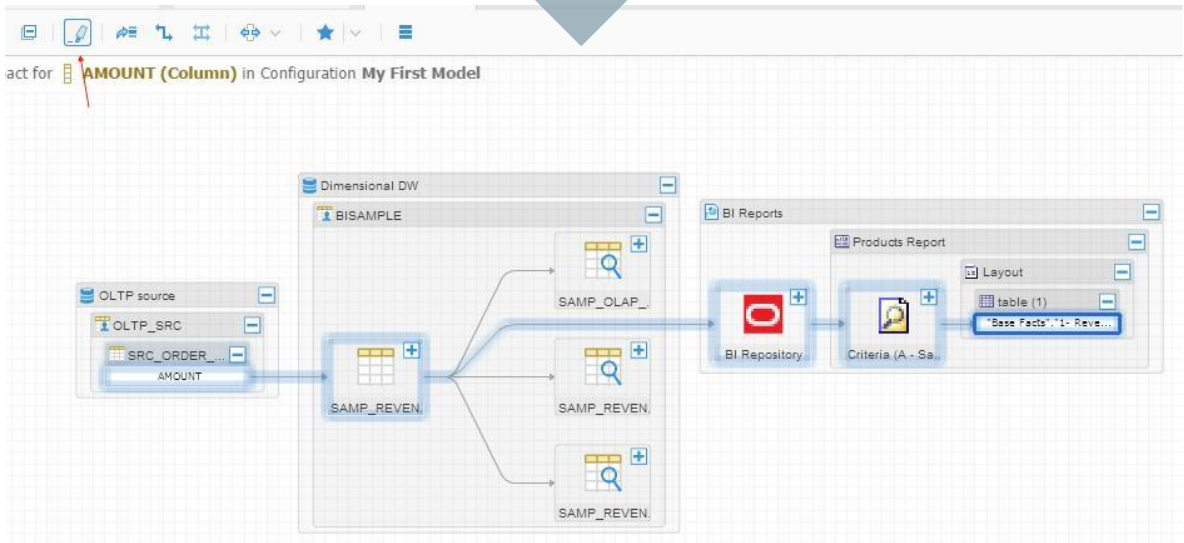




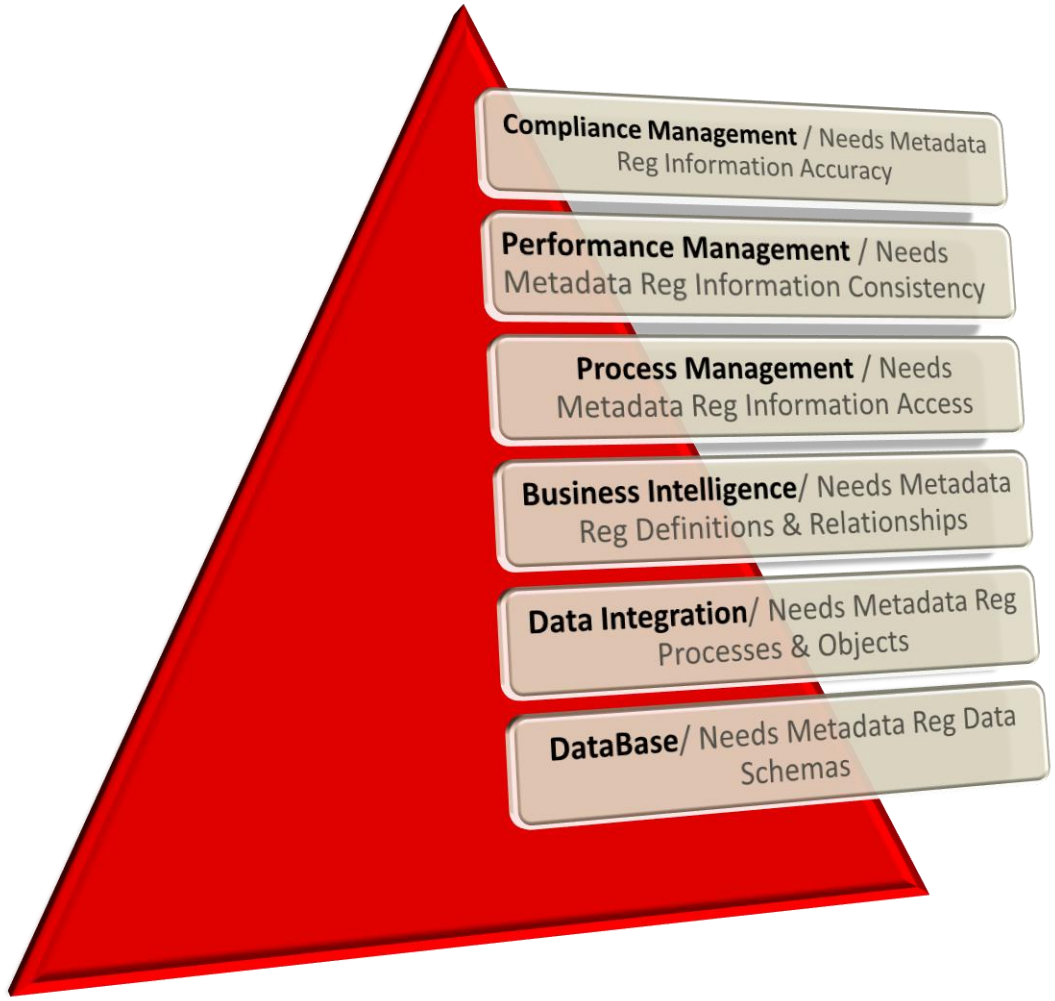




Trace Data Impact
Trace data impact.

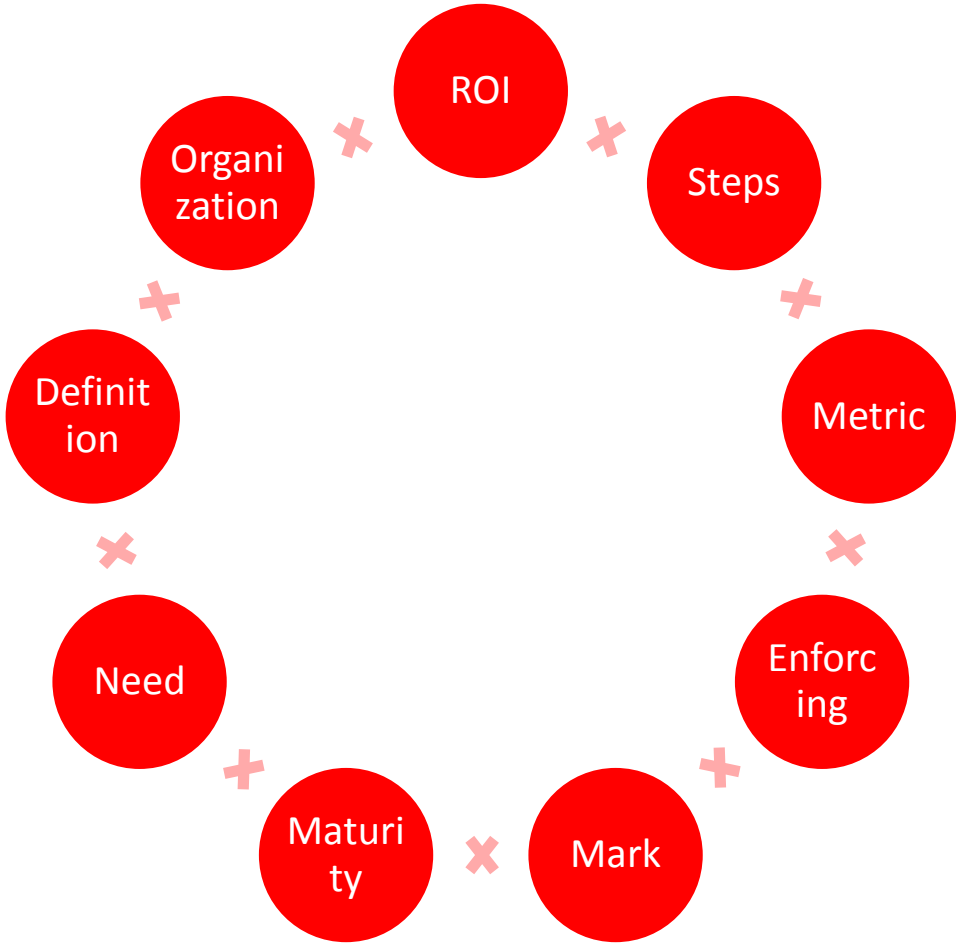


Metadata Requirements Stack

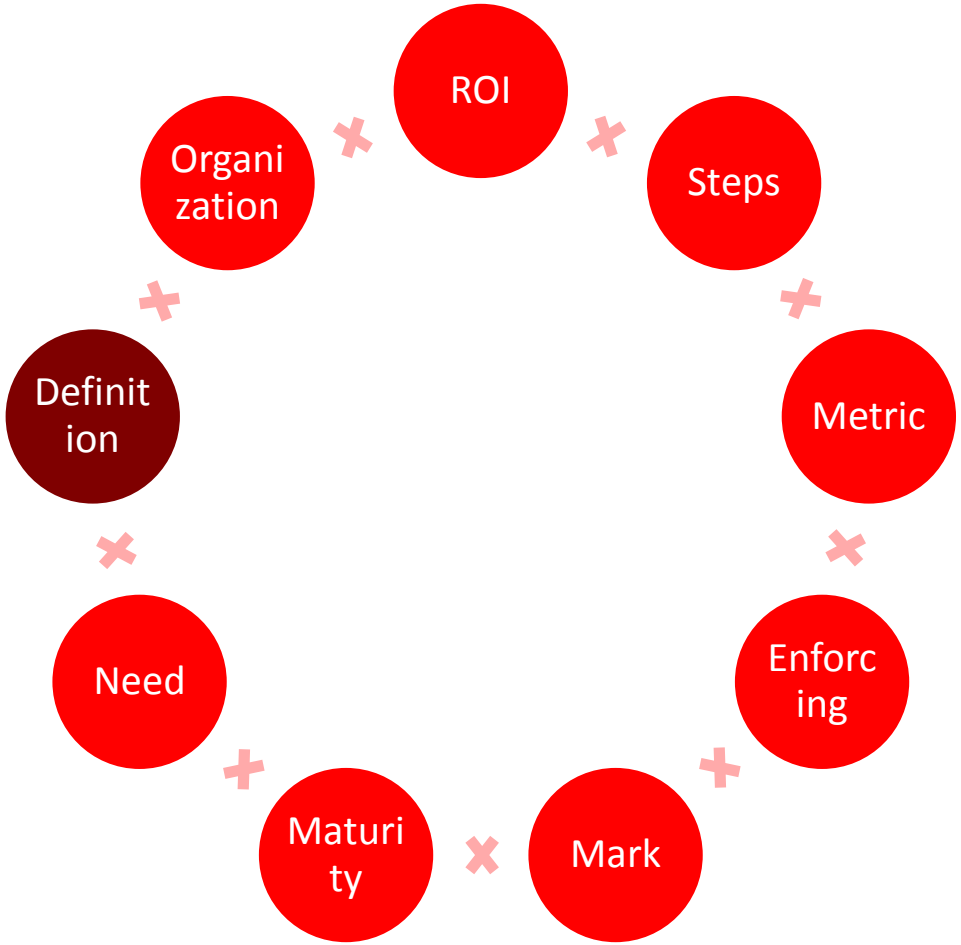


Data Governance

Data Governance Best Practices



Data Governance Best Practices

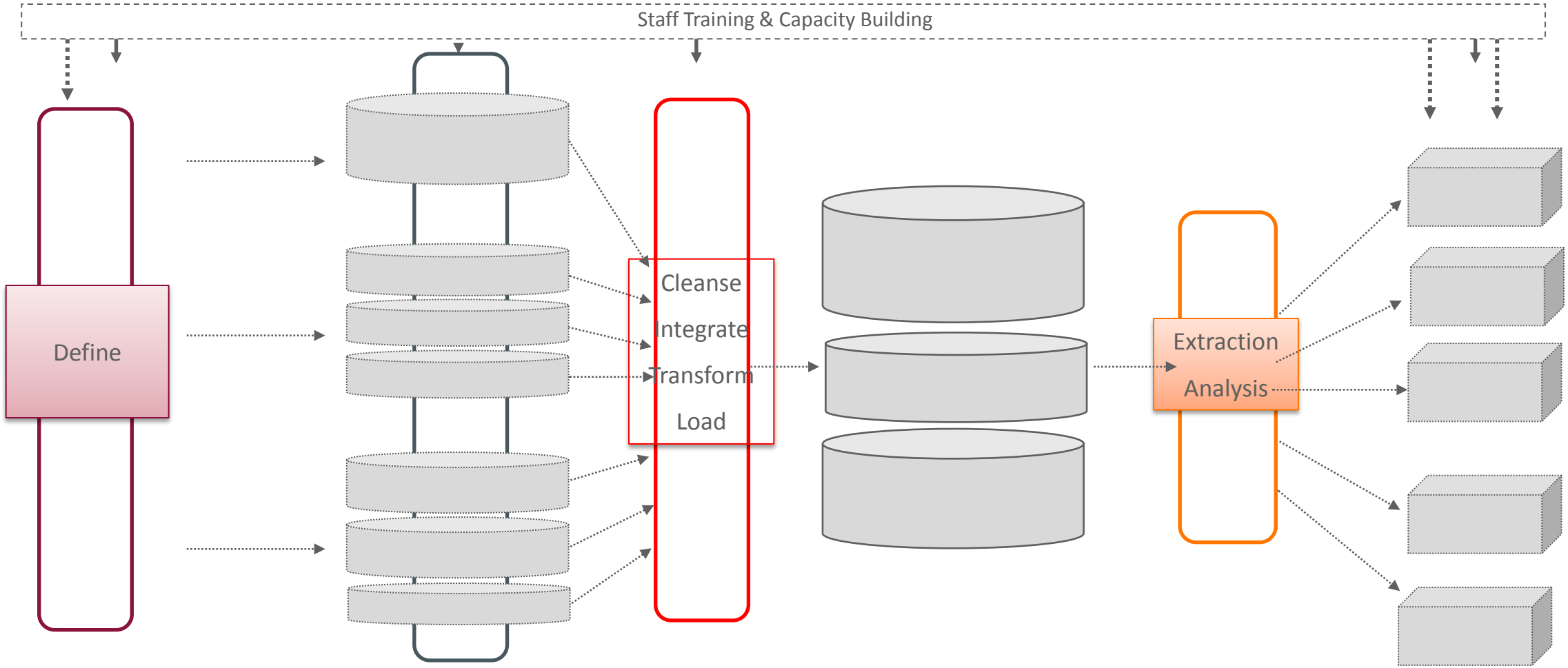


A (Wikipedia) Definition of Data Governance

"Data governance encompasses the people, processes, and information technology required to create a consistent and proper handling of an organization's data across the business enterprise."

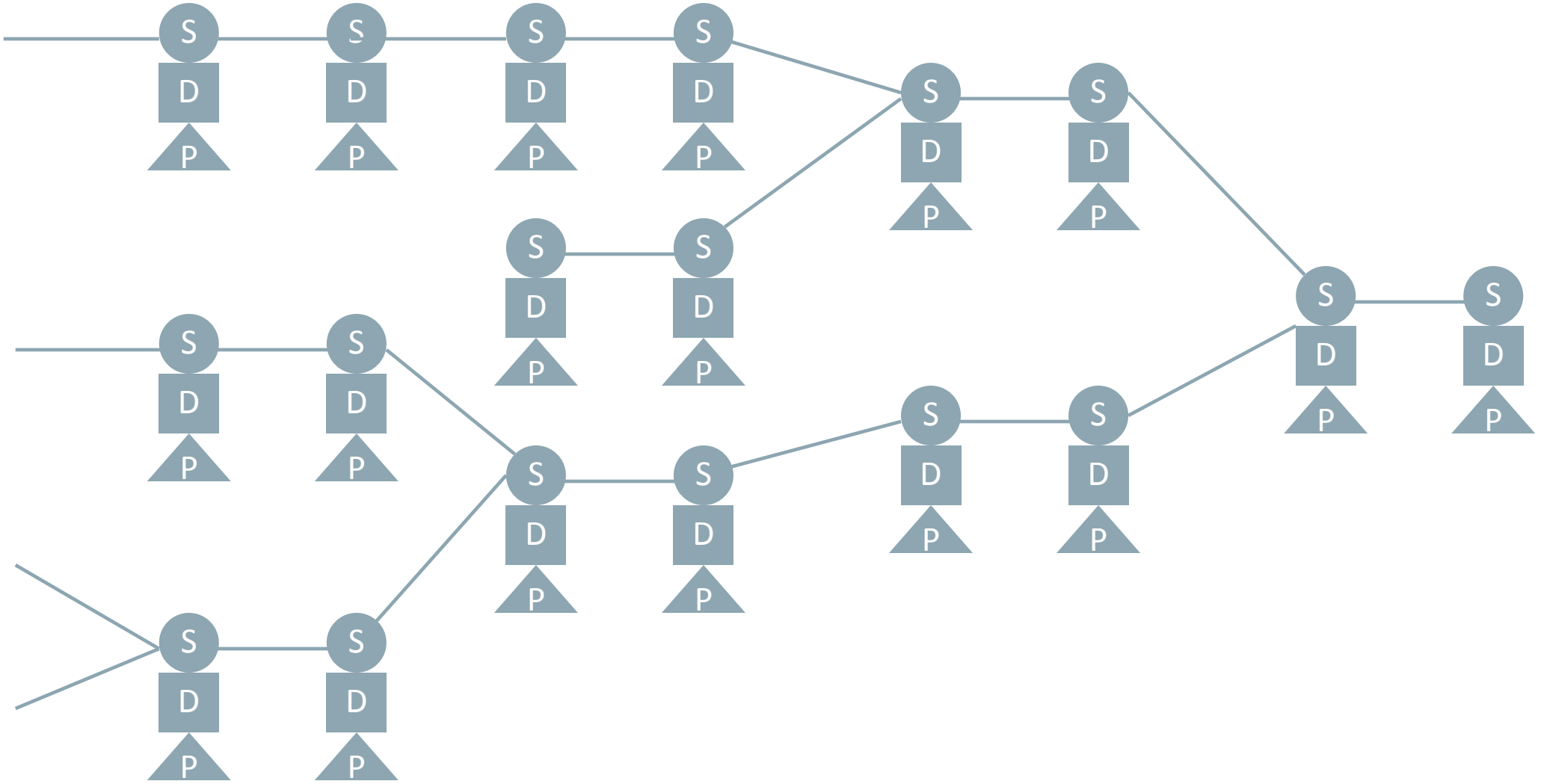
Wikipedia

Data Definition, Production And Usage

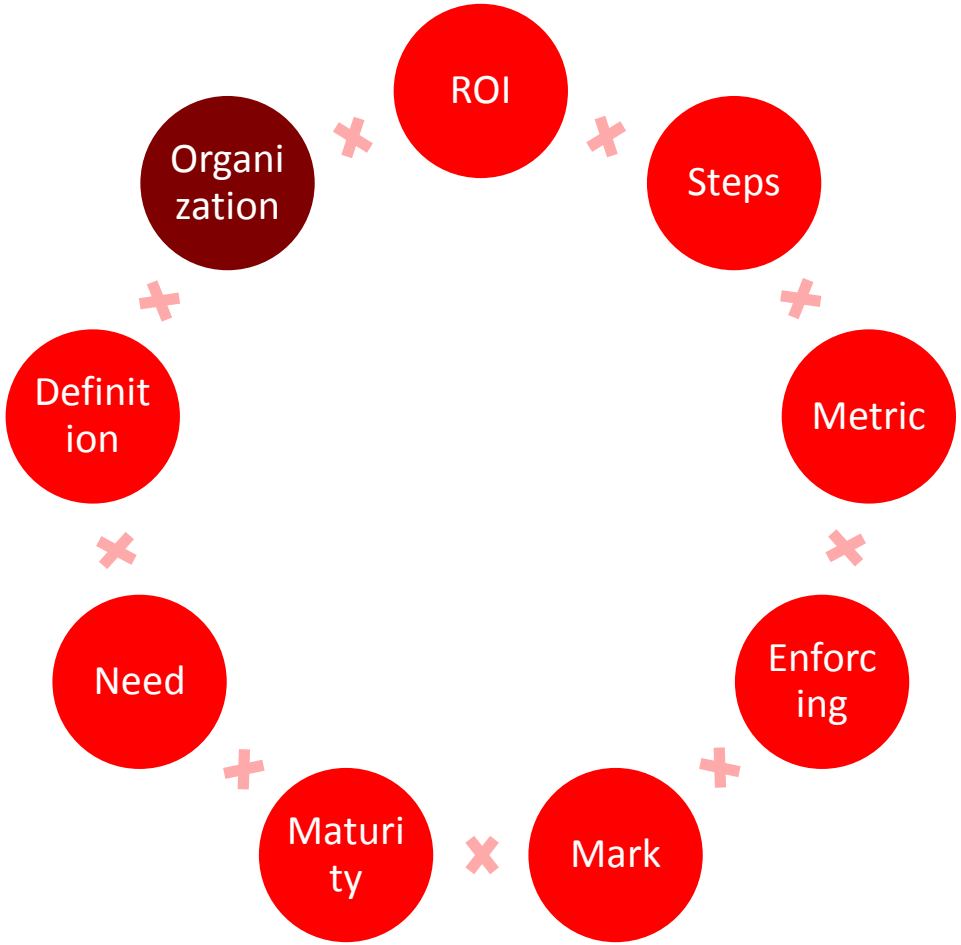


3 5 10 8

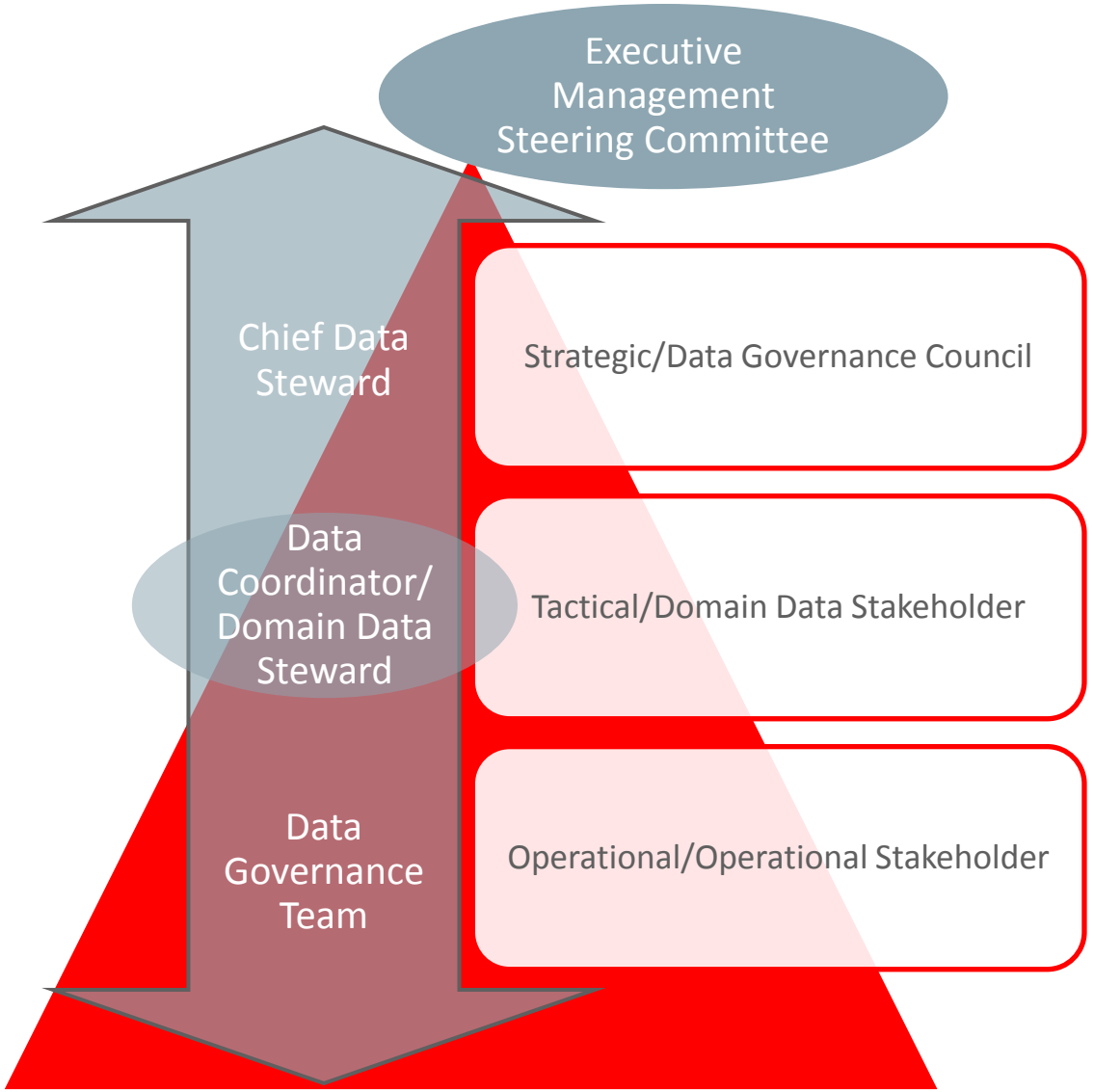
Documenting Back



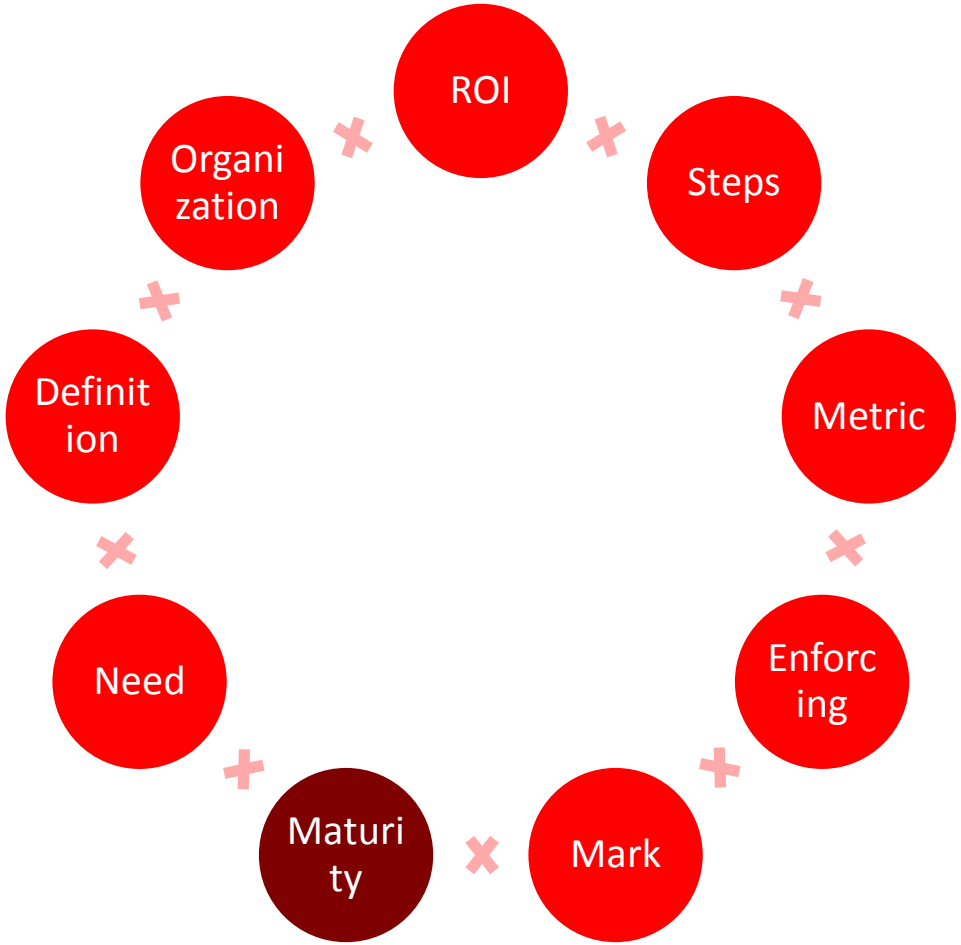
Data Governance Best Practices



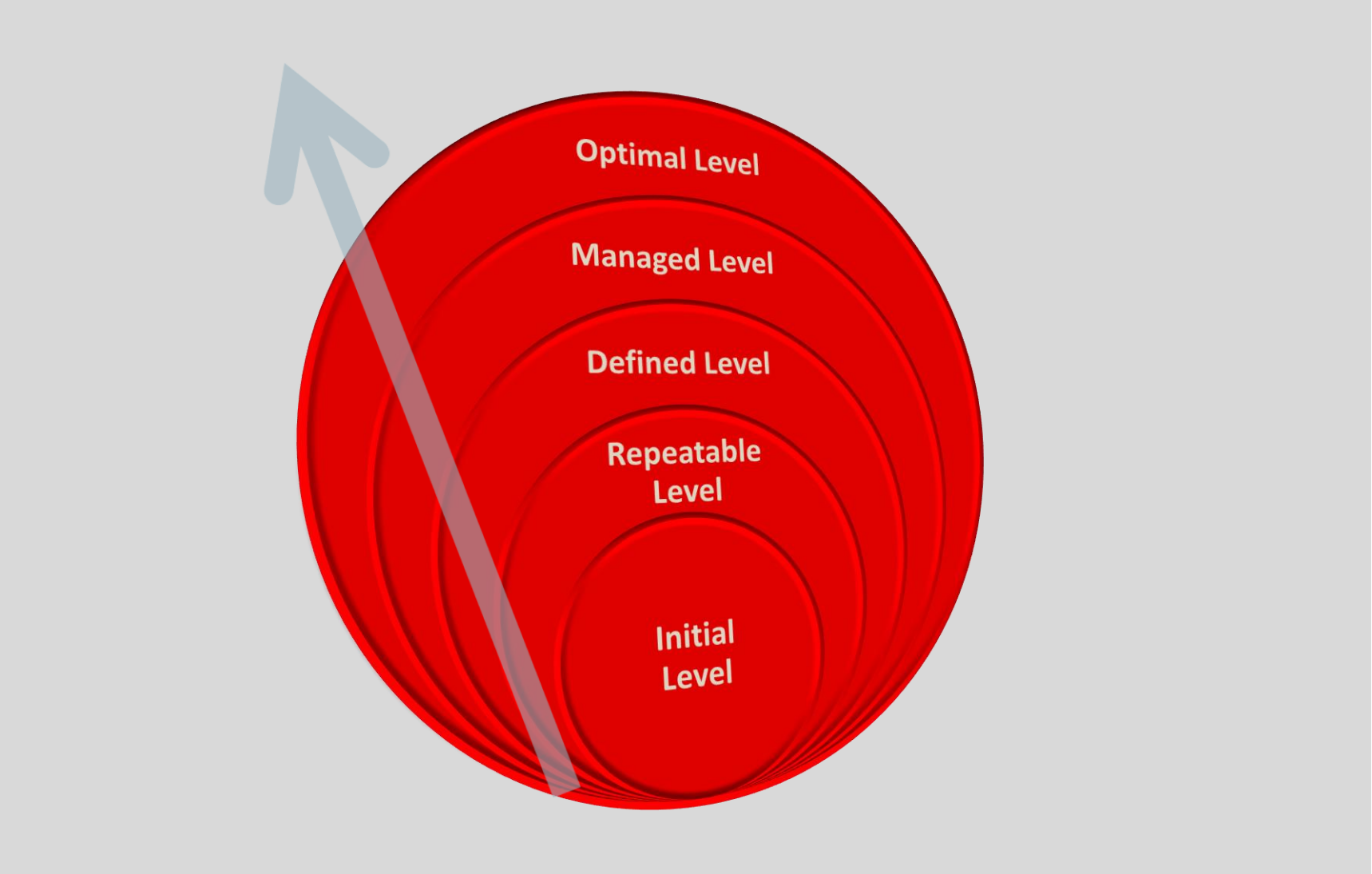
Data Governance Teams



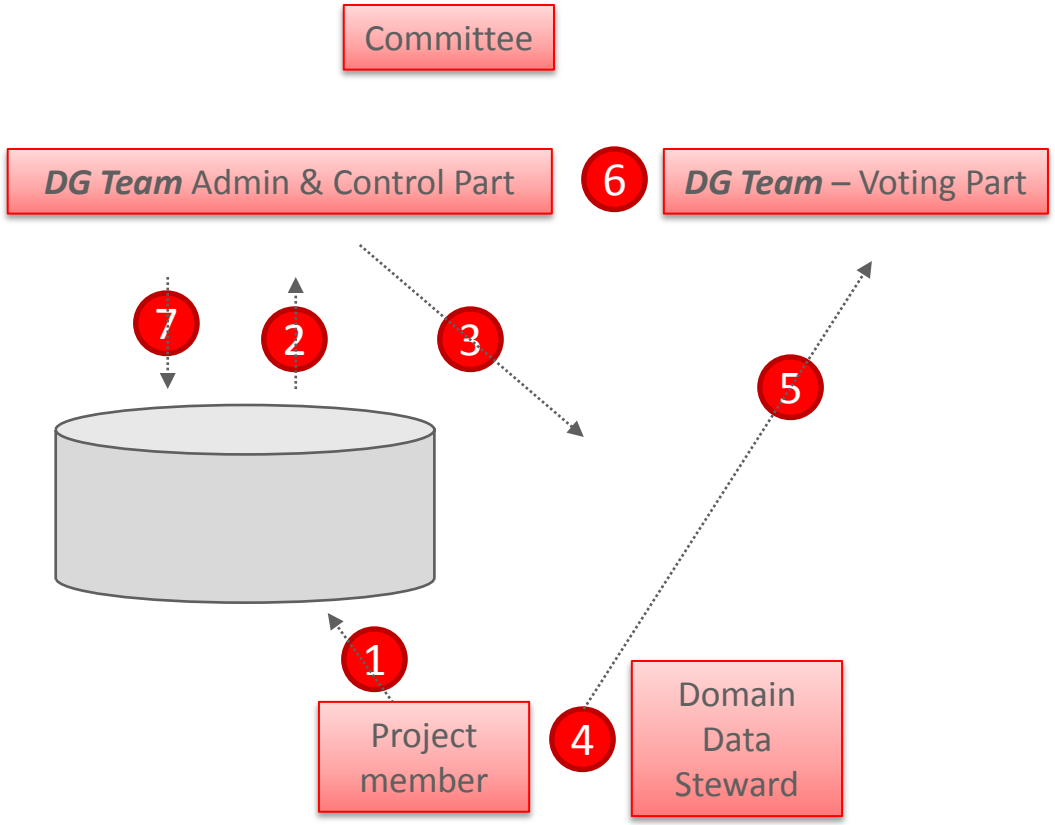
Data Governance Best Practices



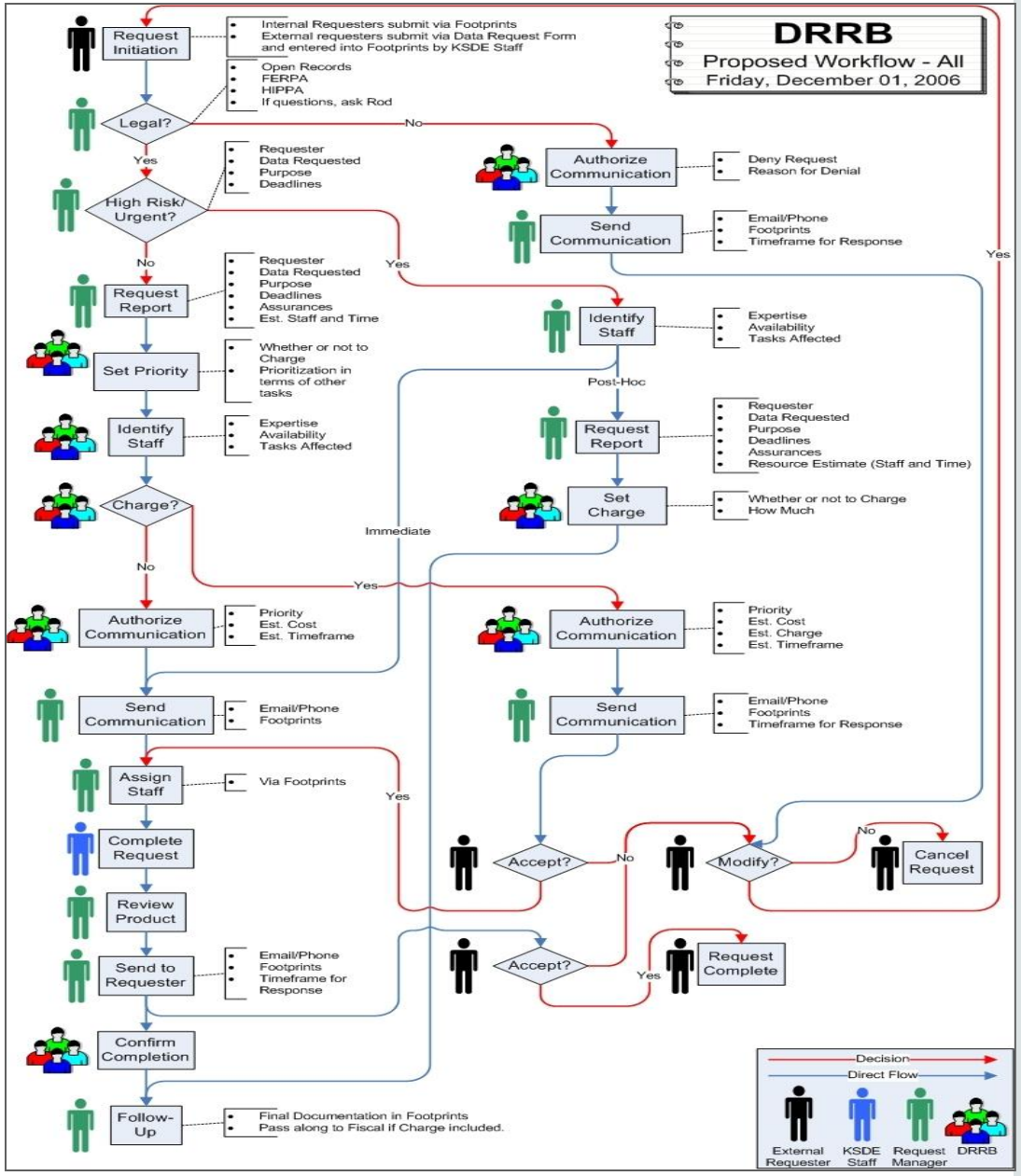
Oracle Data Governance Maturity Levels



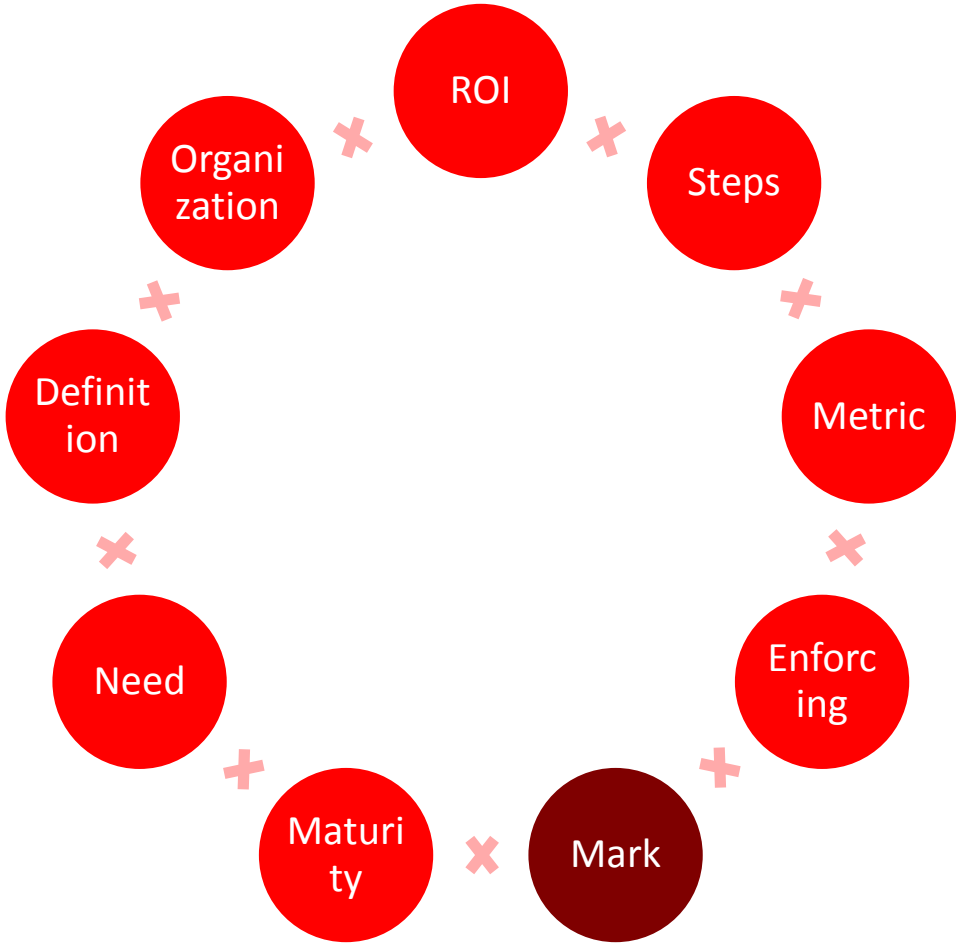
Steps In Workflow



1. Project team submits a term to OEMM for registration
2. Admin receives and monitors this request info
3. The submitted item is routed to Data Stewards
4. Data Stewards work with the project teams to define terms and definitions
5. After definition – submit for approval to the DG team
6. Term is approved by the DG voting members
7. Term is certified for use in OEMM registry & updated in OEMM Business Glossary



Data Governance Best Practices

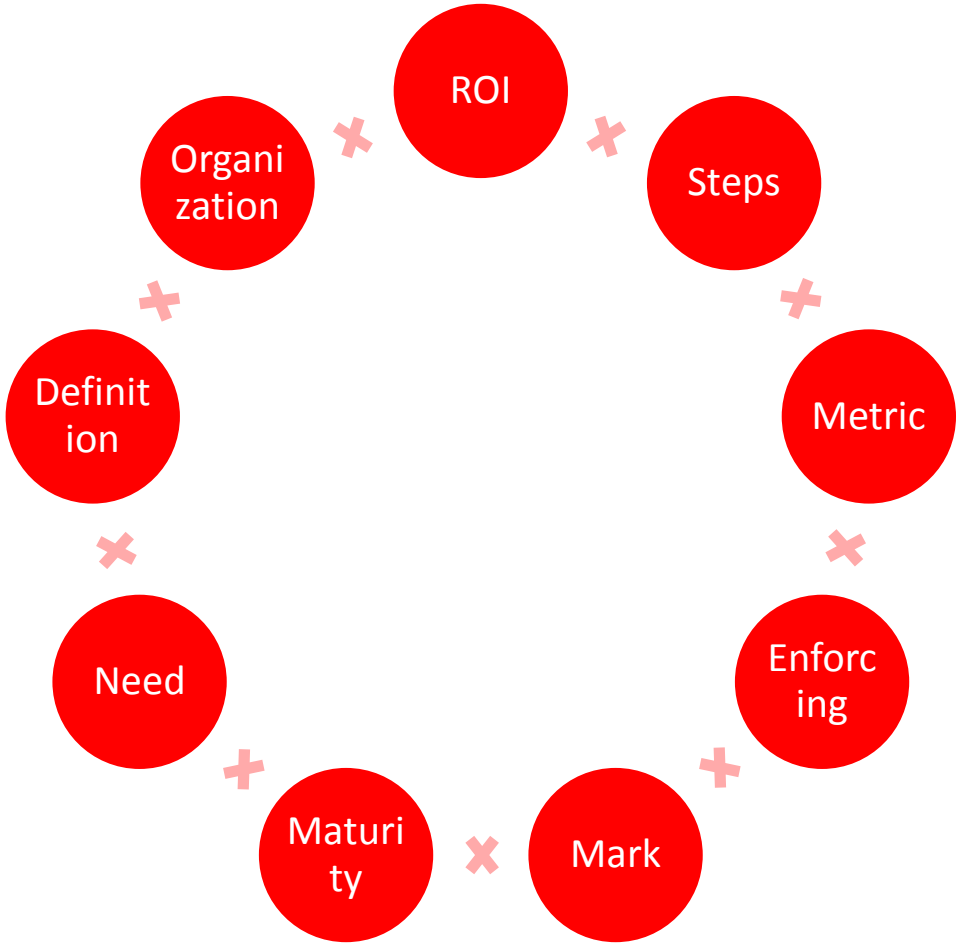


Data Governance Test

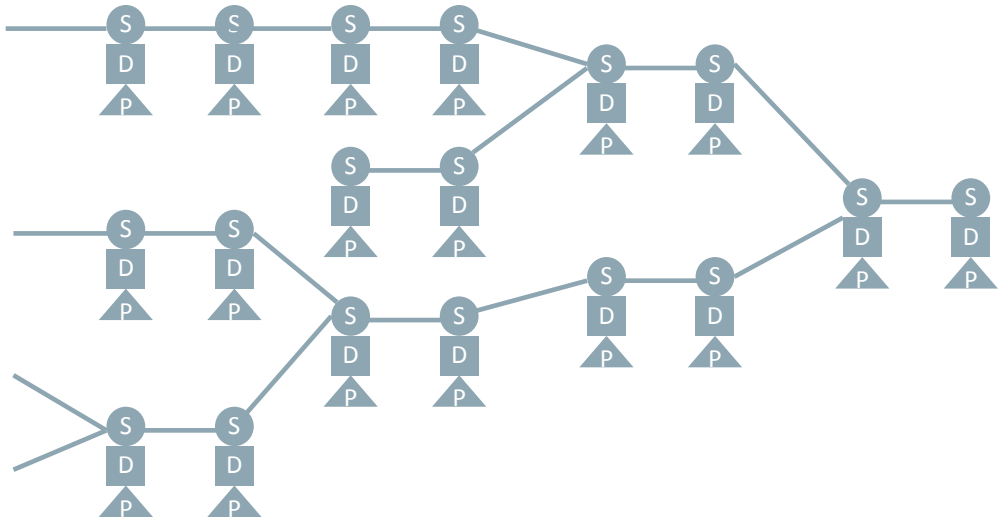
1.	**Risk Management:** We manage the risks associated with our data. My organization understands the need to quickly adjust to the risks associated with data, and many of these rules are coming from outside of the organization. We have a person, a staff of people, or council (or all of the above) that focuses on understanding all levels of risk around the management of data. The person/staff/council regularly communicates information about data risk so that everybody understands risky behavior versus safe behavior in how we handle our data.	1
2.	**Data Compliance & Regulatory Control**: As an organization, we pay a great deal of attention to compliance and regulatory concerns around the data we collect, use, and share as part of making decisions and doing business. Somebody has the responsibility for documenting and communicating the rules to all individuals in the organization who handle these data. When we're audited, we can clearly demonstrate to the auditors that we follow the rules around the data.	3
3.	**Information Security & Data Classification**: As an organization, we pay a great deal of attention to information security for all structured and unstructured data. We have an information security policy and/or something similar, e.g. guidelines, mandates. We feel comfortable with our ability to communicate, differentiate, and manage according to the rules associated with highly confidential data, internal-use data, and public data. People who share data in our organization also share the documented rules about that data, and we don't believe that information security is a concern.	4
4.	**Metadata Management**: We have metadata for the most important data we manage. My organization knows what data we have, where that data resides, and how that data is defined, produced, and used in shared databases and on people's desktops. The information we have about our most important data is available to anybody who needs it. Just as important, we have identified and engaged people who have formal responsibility for the definition, production, and usage of metadata.	5
5.	**Data Quality Management:** Our organization continually focuses on data quality. We have formal means for recording data quality issues, and we have proactive and reactive methods to find issues and address them when we find them. And we have people responsible for managing the issue logs, putting values to the issues and prioritizing the issues. Most important, we have a clear understanding about the business standards for core pieces of data that make it easier to differentiate high quality from low quality data.	1
6.	**Business Intelligence and Data Integration:** We have a data warehousing environment that takes full advantage of the data therein and is used to its fullest capability. This means that people have easy access to the data, they understand the data, and they help us to continually improve the quality of the data. We recognize that data governance plays an important role in the success or failure of our data warehousing initiative on all sides of the data integration equation. We understand that data integration is a difficult discipline. But since we govern the data well on both sides—source and target—we feel comfortable with the effective nature of our business intelligence program.	2
7.	**Master Data Management:** Our organization recognizes that master data management (MDM) is one of the most effective and most important data disciplines talked about today. We've identified people to manage our MDM initiative(s) and have started to identify the enabling technologies that will help us manage and share our master and reference data. When we populate our MDM environment, the discipline is there to manage the decision-making around the master data resource, the metadata component, and the communications and accessibility to the master data. We are positioned well to complete the master data initiative within budget and on schedule.	3
8.	**Data Governance & Data Stewardship:** Last but not least, we have a data governance program that clearly defines roles and responsibilities at the operational, tactical, strategic, and support levels. Our program focuses on leveraging the existing knowledge of the data that lies within our data stewards. The approach we've taken has been embraced by our leadership, stewards, business, and the technology individuals and it addresses the governance of data in a proactive and reactive sense. Our data governance program is a primary contributor to our success in all of the disciplines listed in this test.	2

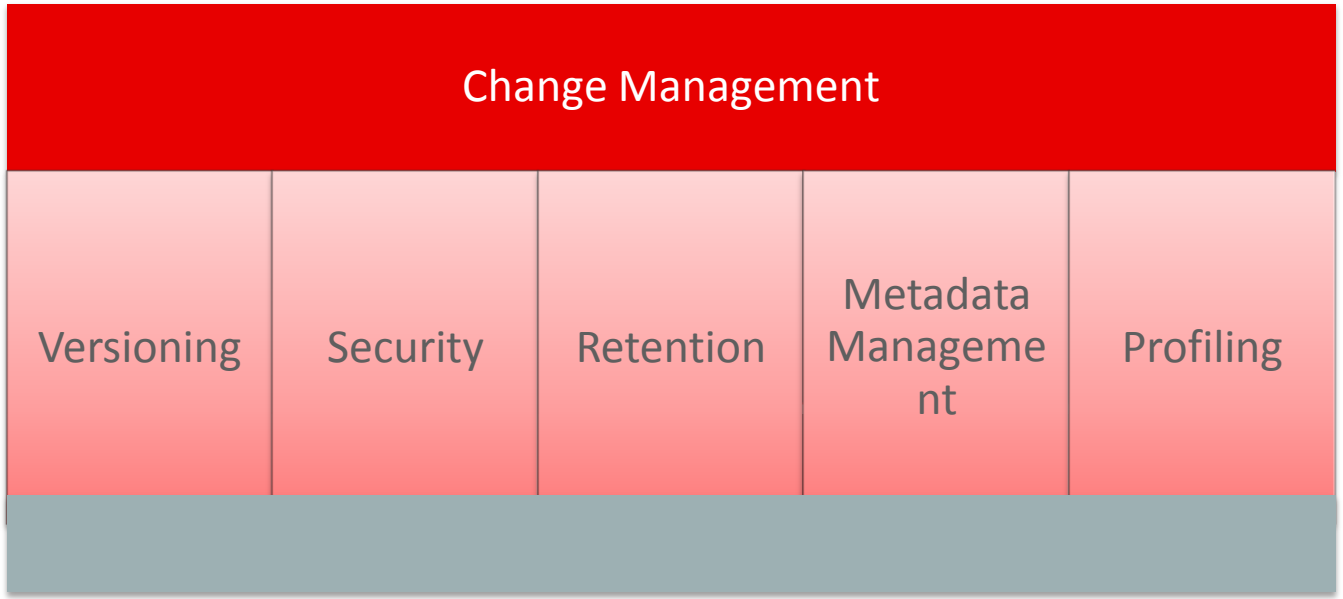
Source : Safari Books Online

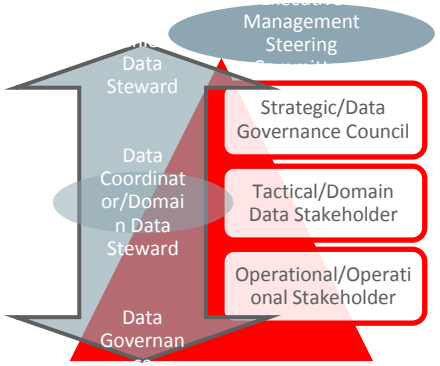
Data Governance Best Practices

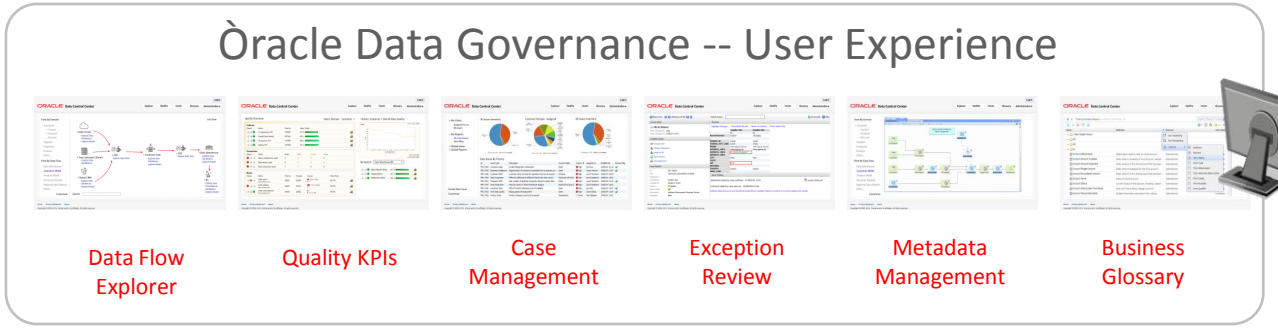


Data Governance Best Practices

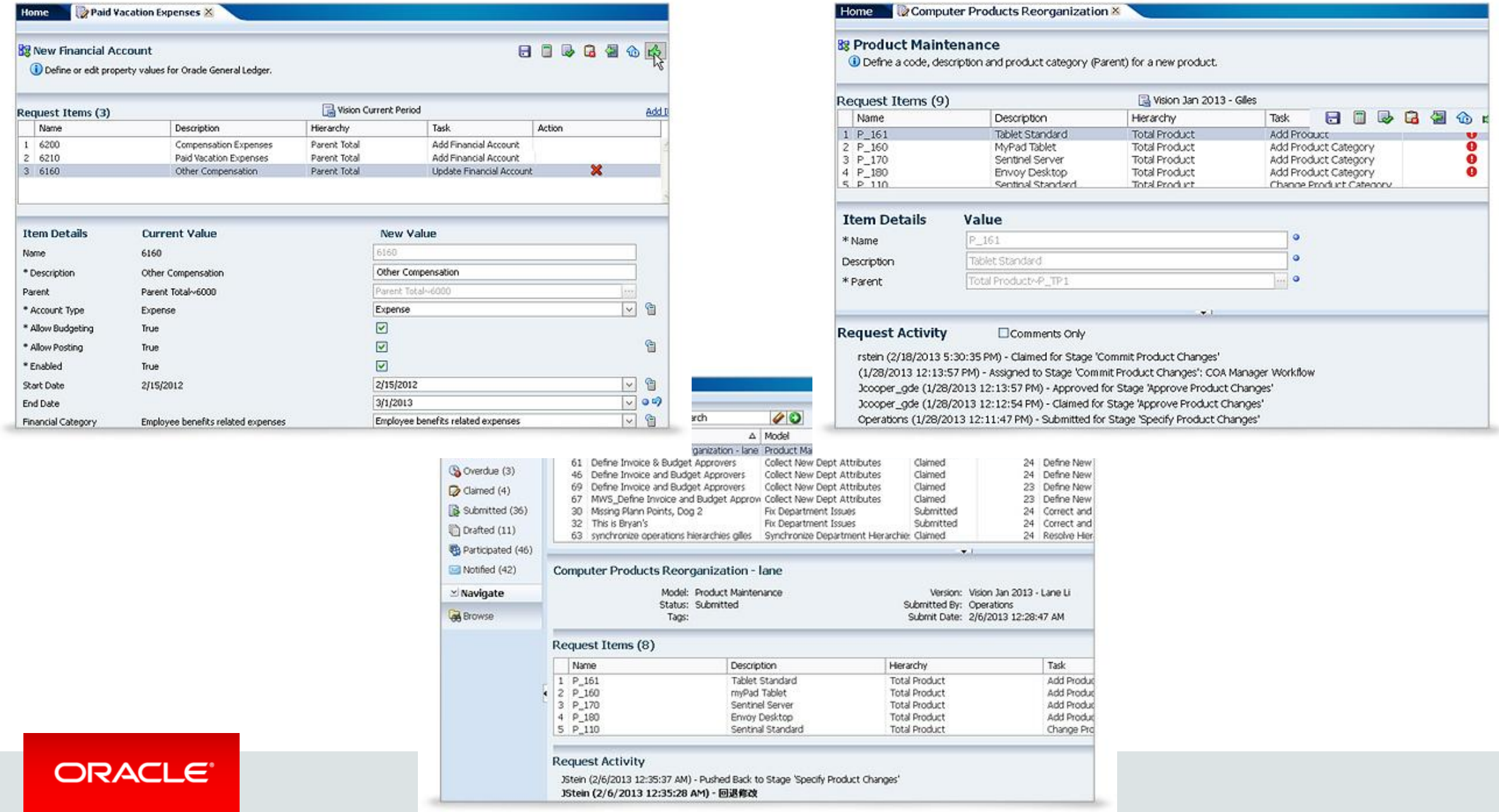




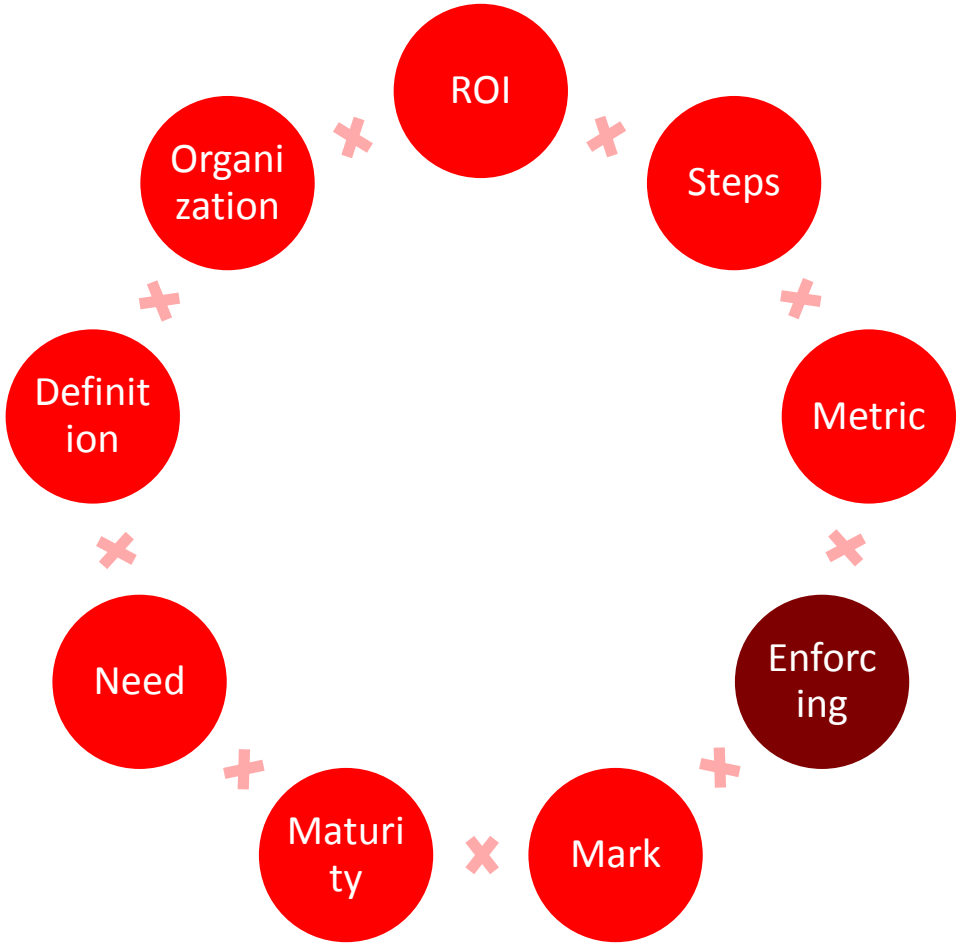




Third Step In Oracle Data Governance - DRM



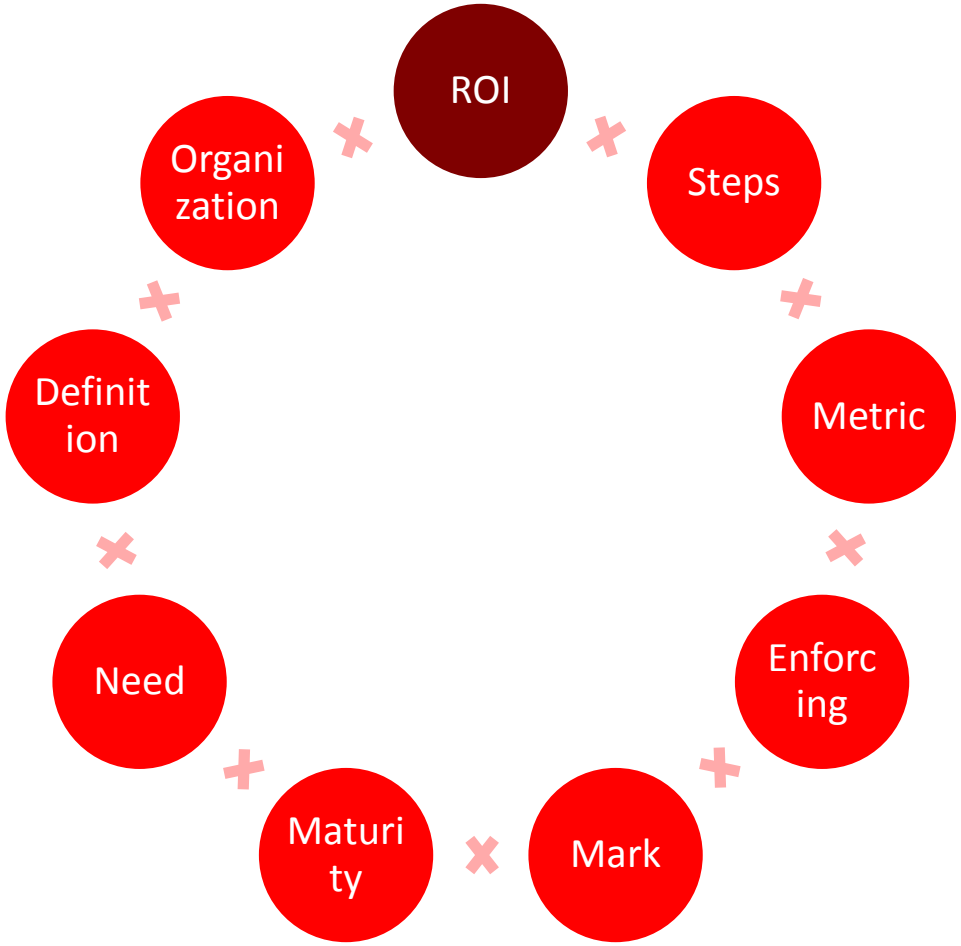
Data Governance Best Practices



Enforcing Data Governance

		DG Council	DG Team	Domain DG Team	Operational DG Team
Identification& Documentation	Obtain the position job description from HR	I	R/A	R	R
	Identify the data touch points associated with position	I	R/A	R	R
	Identify the rules associated with data		R/A	R	R
Ratification	Attach accountabilities to data touch points	I	R/A	C	C
	Communicate accountabilities associated with data	I	R/A	I	I
Implementation& Control	Enforce accountabilities associated with data		R/A	C	R
	Measure accountabilities associated with data	I	R/A	C	I
	Measure accountabilities associated with data	I	R/A	C	I

Data Governance Best Practices



ROI – 5 Areas To Demonstrate Value

(1) Program
Strategy & Direction
Structure
Scope
Prioritization
Policy
Roles & Responsibilities
Authority & Accountability
Issue Resolution
Business Alignment
Power To Question
Transparency
Weight & Balances
CoE
Agility To Respond

(2) Data Management Operations
Data Quality
Data Modeling
Metadata Management
Data Integration
Data Lineage
Security & Privacy
Data Standards
Business Rules
Business Process
Stewardship
Monitoring&Metrics
Reporting
Workflow Management

(3) Projects
MDM
BI
CRM
CDI
DW
EDW
DM
ERP
360 View
Data Migration
Data Integration

(4) Business Operations
Sales & Marketing
Customer Service
Risk & Compliance
Order To Cash
Supply Chain
Taxes & Fees
Social & Human Services
Mail & Transport
Asset Management
Business Analytics & Research

(5) Organization Strategy & Policy
Customer Definition
Merger & Acquisition
Planning & Budget
Business Partner Integration
Data Enrichment
Security & Privacy

ORACLE

Oracle Data Governance ROI

Increase revenue / value of assets

❑Improve the value of the company to those who would acquire it
❑Utilize information assets to make new sales, Better understand customers, Better understand product (and other) hierarchies

Reduce costs

❑Reduce duplicate data management processes (example: costs of data modeling, data administration, data quality)
❑Reduce errors and associated costs (in software development, report development, information interpretation) due to lack of understanding of data or poor quality data

Support Compliance While Reducing Costs

❑Avoid audit fees due to lack of confidence in “authoritative data”
❑Reduce costs of pre-audit testing

Support Impact Analysis

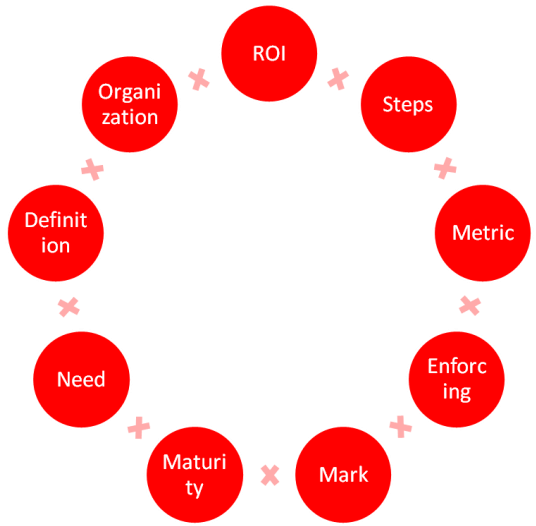
❑Provide a capability to assess cross-functional impacts of data-related decisions, to do useful impact analysis
(by providing authoritative business rules, system of record information, and data lineage metadata)
❑Avoid “undoing” work or rendering controls invalid

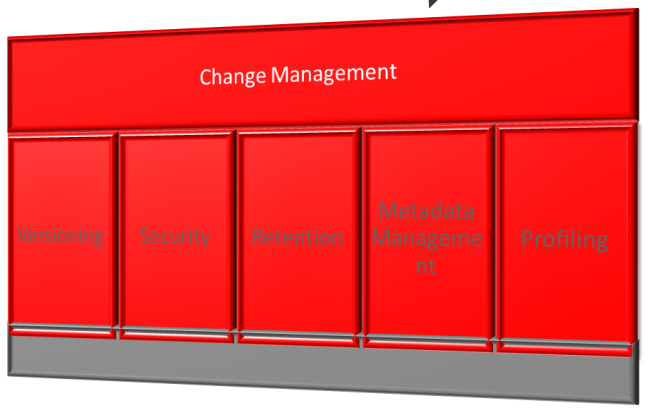
1

Òracle Data Governance -- User Experience

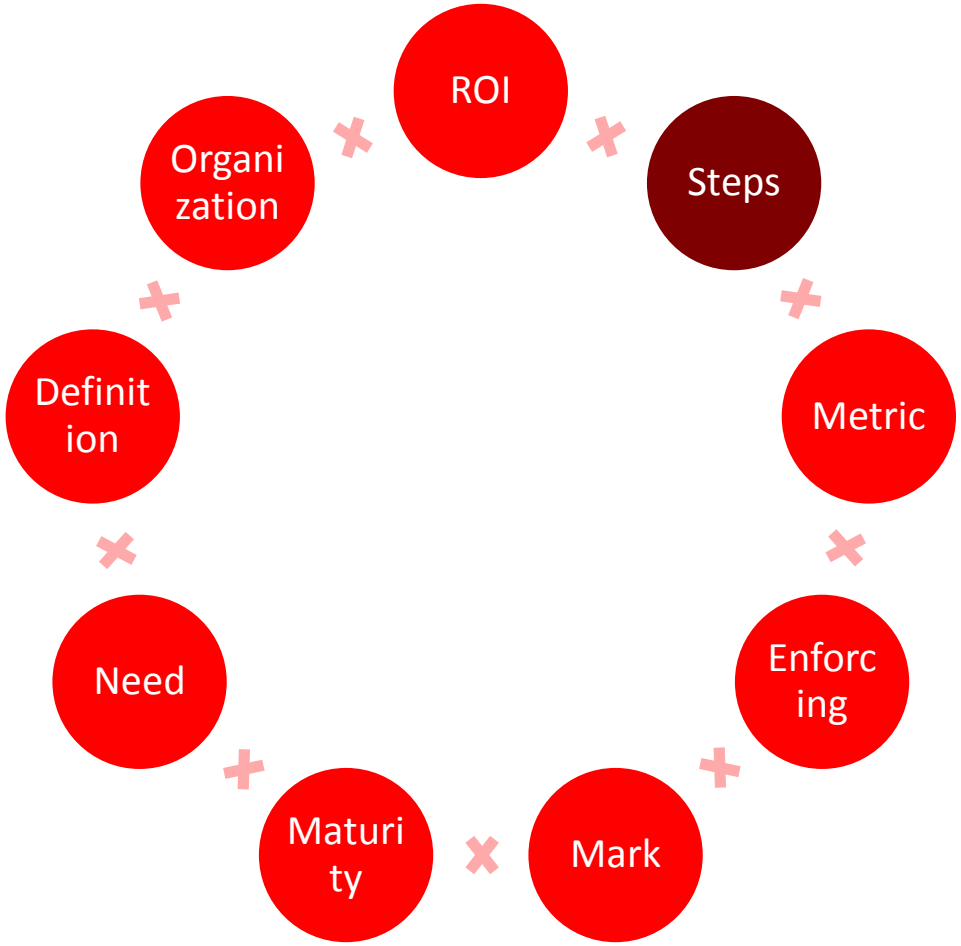
Data Flow Explorer | Quality KPIs | Case Management | Exception Review | Metadata Management | Business Glossary

2

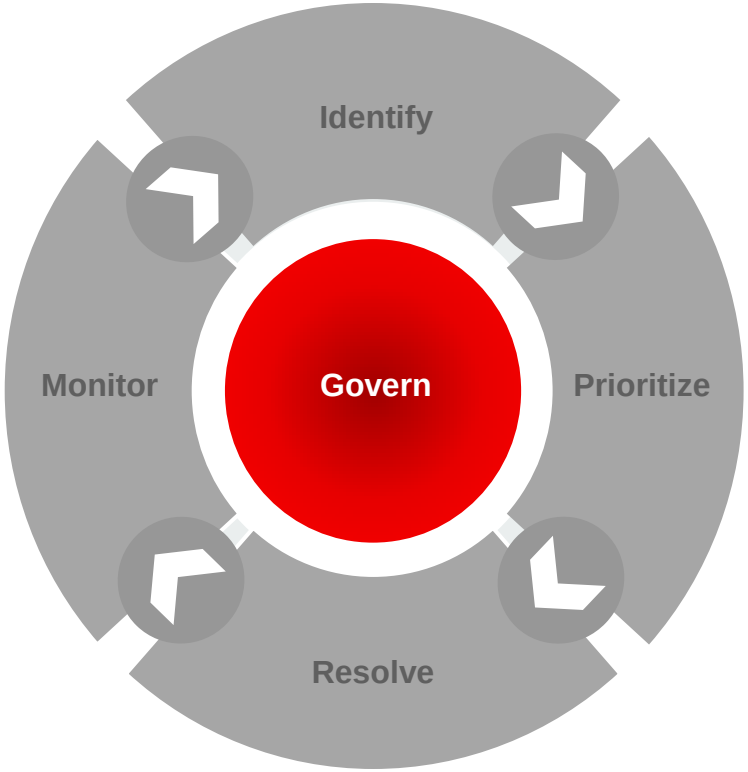


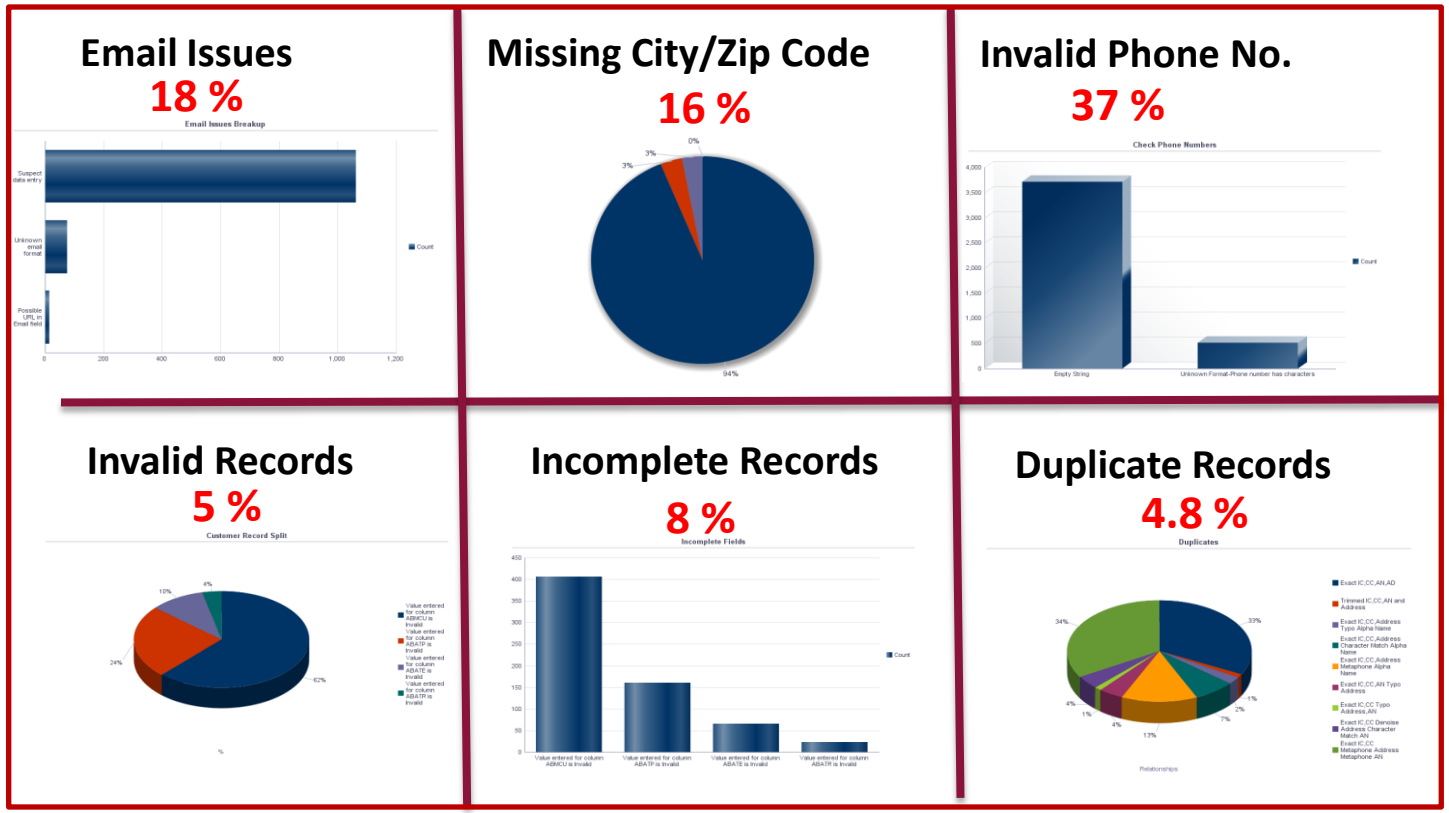


Data Governance Best Practices

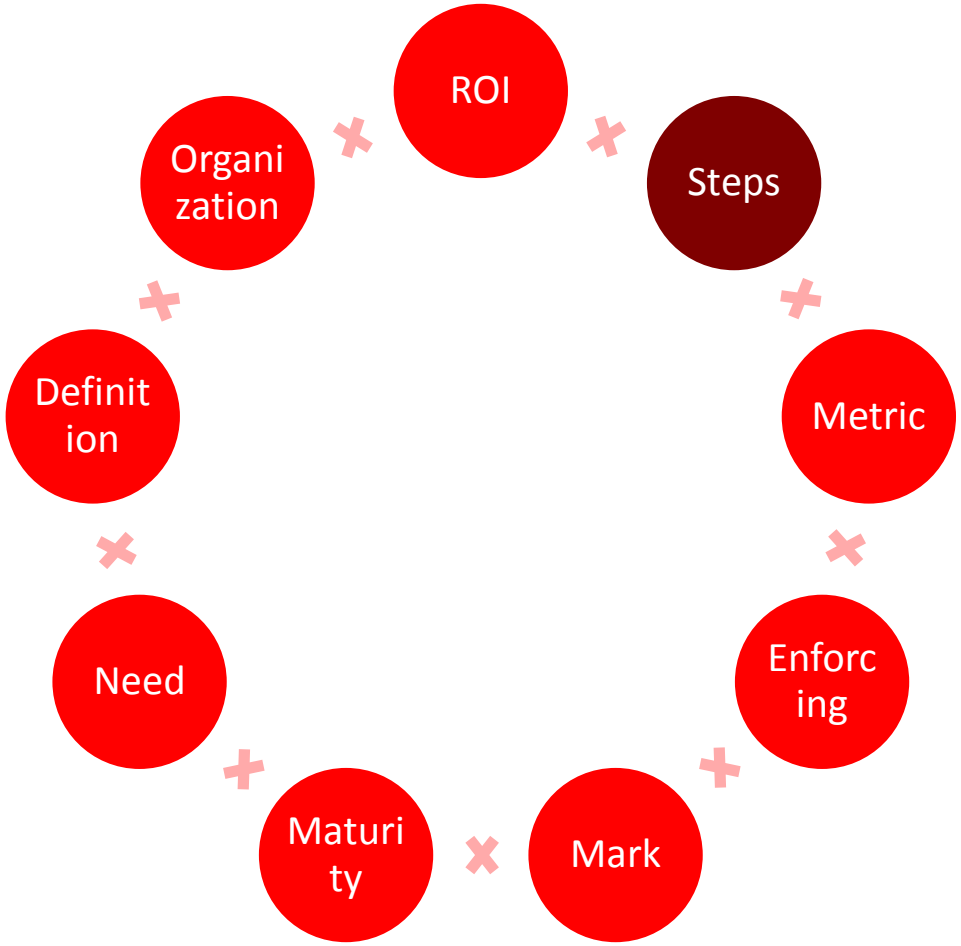


The Data Governance Process





Data Governance Best Practices

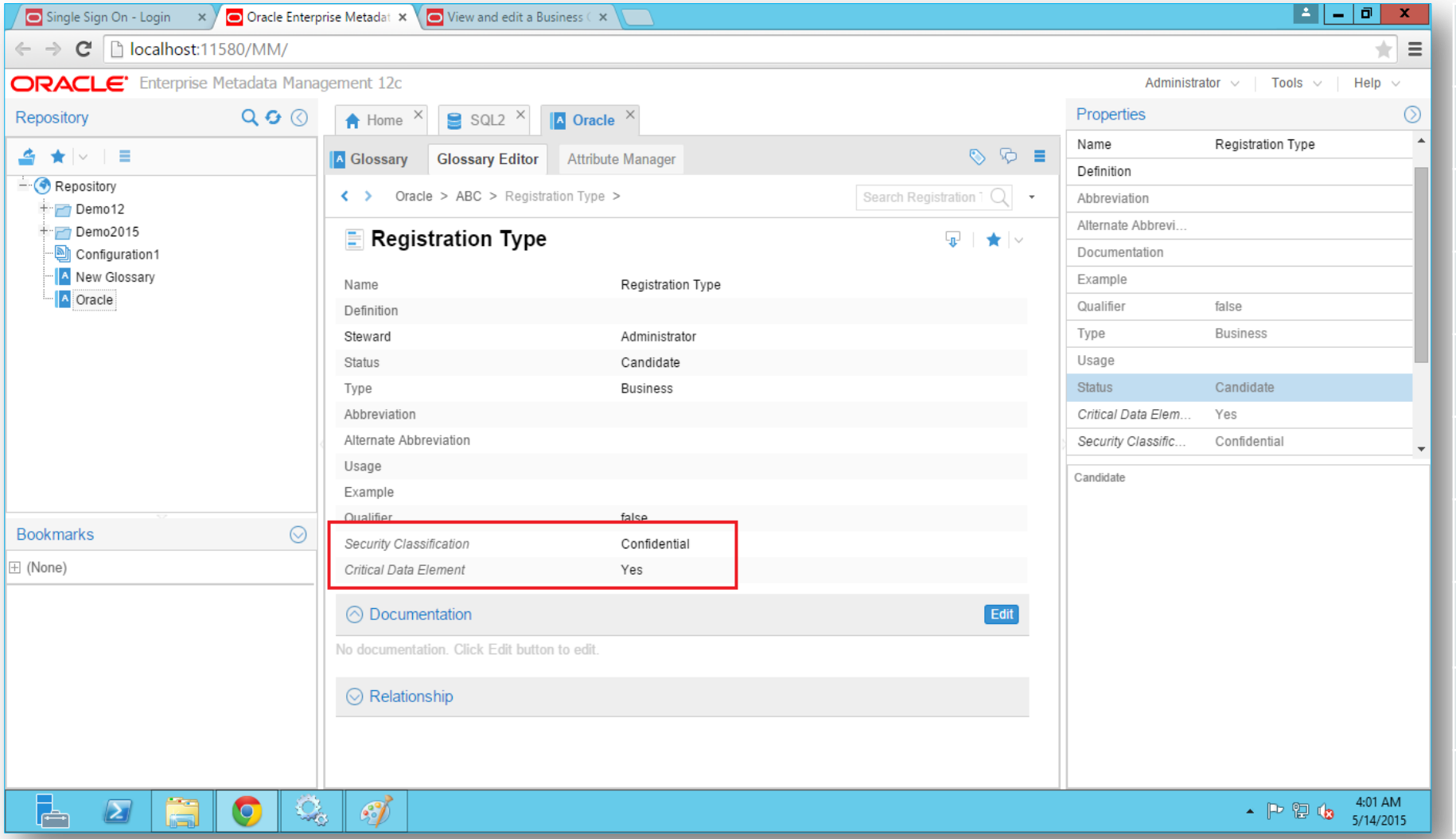


Data Governance Matrix

	IT				Marketing			HR	
1. Customer Data	**System Name**	Domain Steward	Operational Steward	Data Gov Team Member	**Area 1**			**Area 1**	
1.1. Customer Address Data	ERP System								
	MDM DB								
	EDW DB								
1.2. Customer Demographics Data									
	ERP System								
	MDM DB								
	EDW DB								
1.3. Customer Financial Data									
	ERP System								
	MDM DB								
	EDW DB								

Critical Data Elements in OEMM

- **Critical Data Elements**
- Business Glossary
- Stewardship Workflows
- Business Rule Definitions
- Rule Semantics
- Term to Rule Linking
- Allowable Values
- Search and Browse
- Data Flow Lineage
- Model Versioning

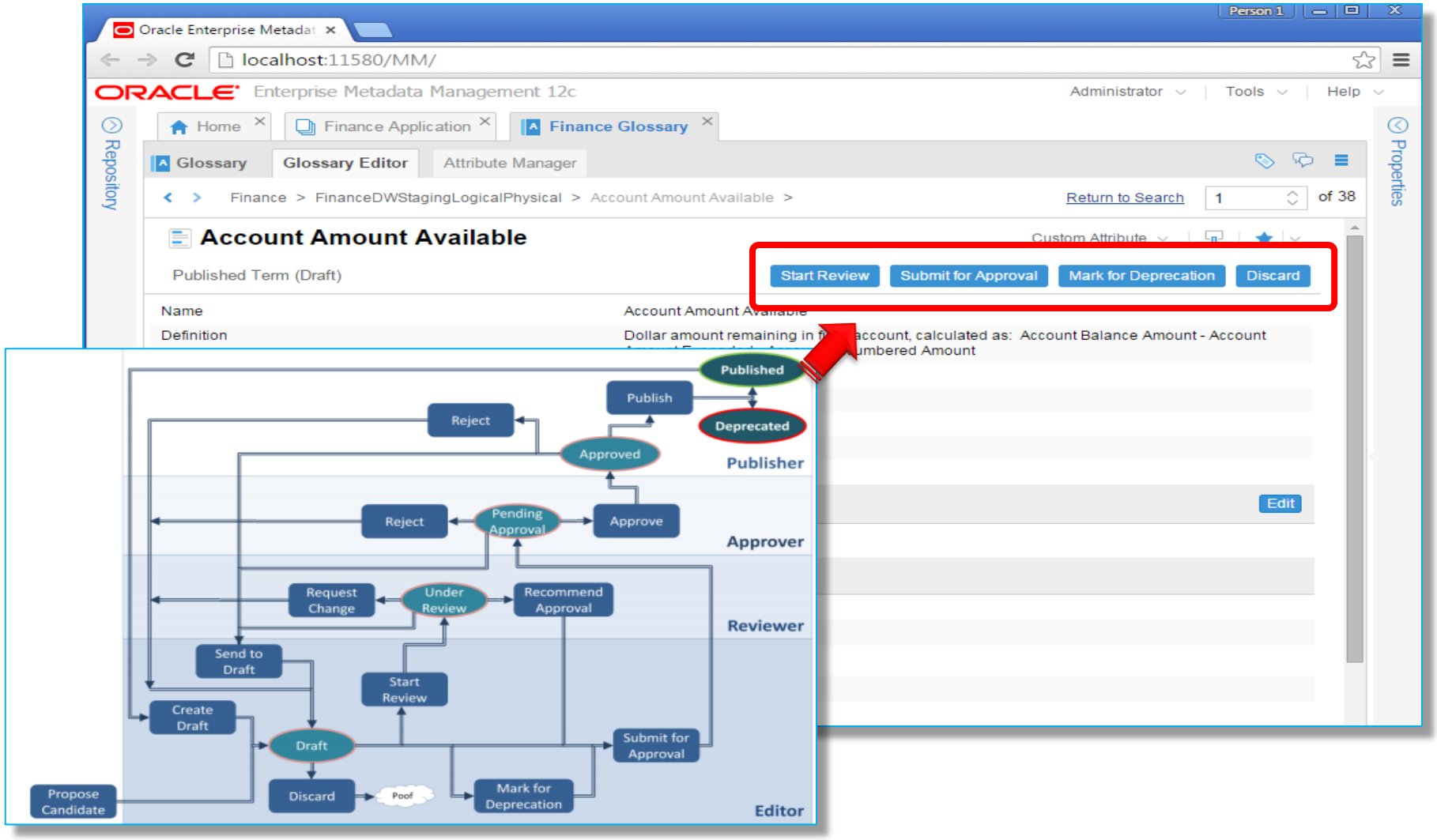


- Model Browser
- Systems Lineage
- Data Profiling
- Data Validation
- Dashboards
- Remediation & Repair
- 3 Key Users
- 3 Kinds of Lineage
- 3rd Party Integrations
- Deployment Architecture

Source : Oracle Product Management and company Information Assets

Business Glossary Stewardship Workflow

- Critical Data Elements
- Business Glossary
- **Stewardship Workflows**
- Business Rule Definitions
- Rule Semantics
- Term to Rule Linking
- Allowable Values
- Search and Browse
- Data Flow Lineage
- Model Versioning

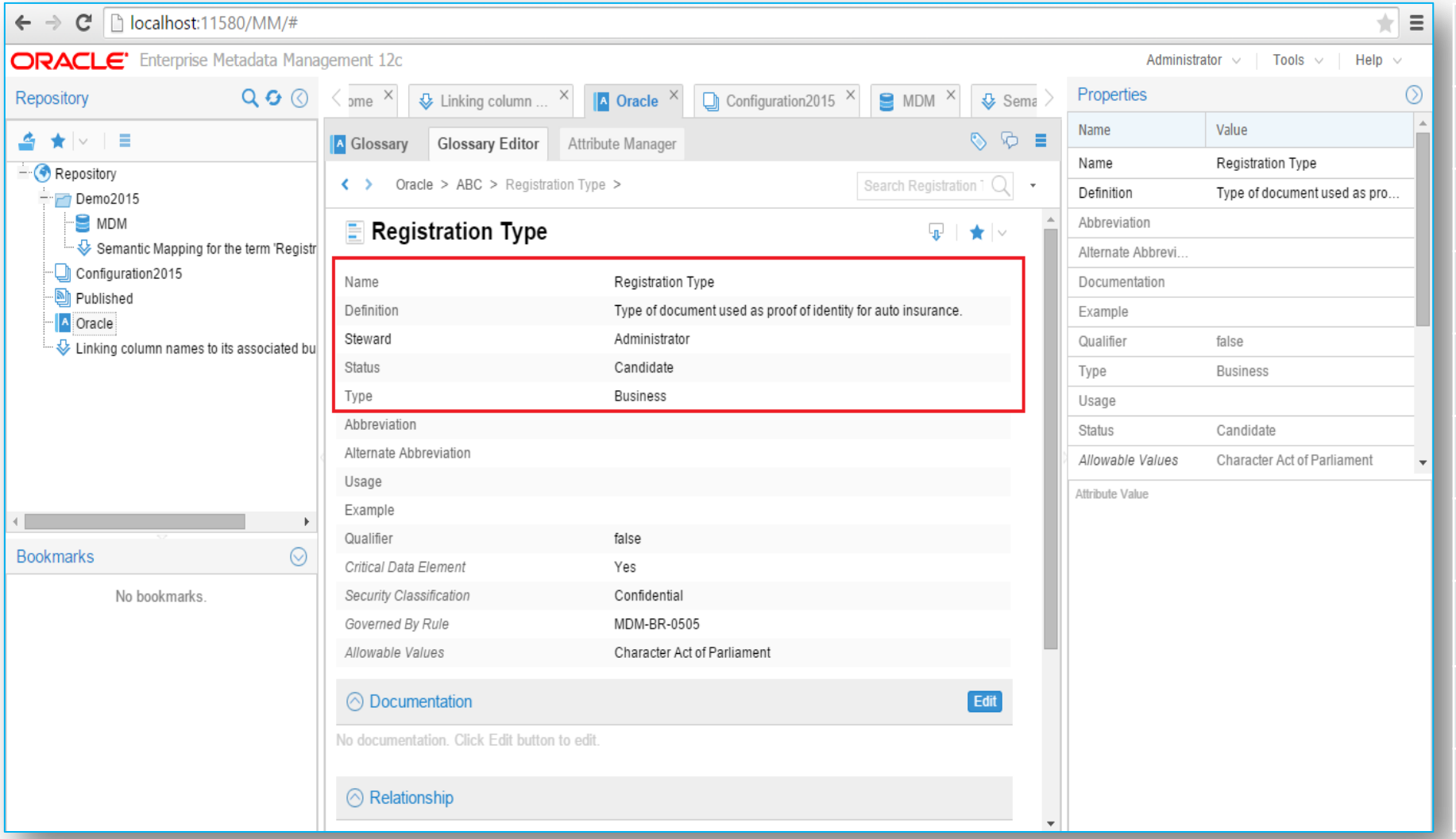


- Model Browser
- Systems Lineage
- Data Profiling
- Data Validation
- Dashboards
- Remediation & Repair
- 3 Key Users
- 3 Kinds of Lineage
- 3rd Party Integrations
- Deployment Architecture

Source : Oracle Product Management and company Information Assets

Business Glossary Manages Definitions

- Critical Data Elements
- **Business Glossary**
- Stewardship Workflows
- Business Rule Definitions
- Rule Semantics
- Term to Rule Linking
- Allowable Values
- Search and Browse
- Data Flow Lineage
- Model Versioning

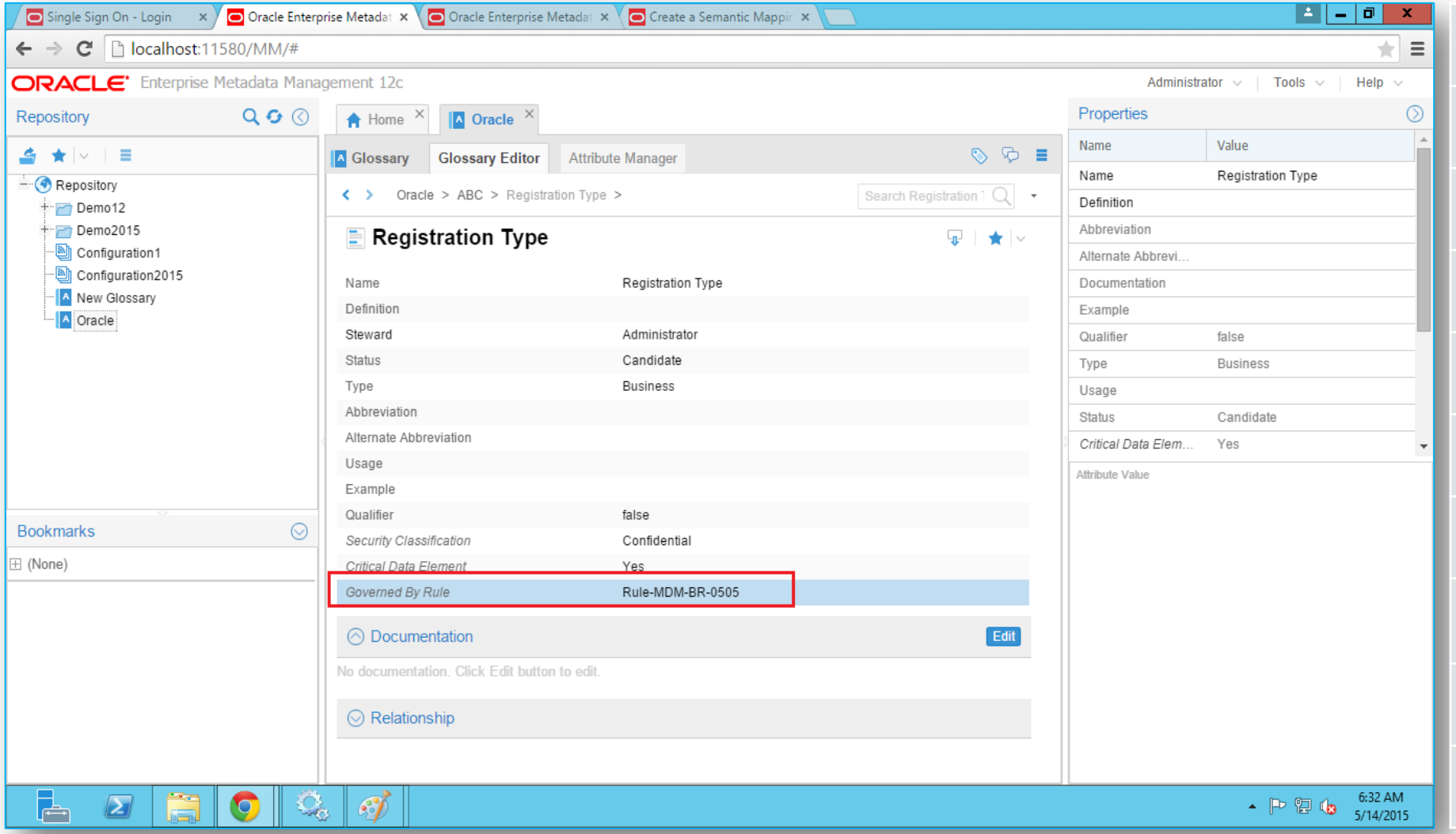


- Model Browser
- Systems Lineage
- Data Profiling
- Data Validation
- Dashboards
- Remediation & Repair
- 3 Key Users
- 3 Kinds of Lineage
- 3rd Party Integrations
- Deployment Architecture

Source : Oracle Product Management and company Information Assets

Mapping of Business Term to Rules

- Critical Data Elements
- Business Glossary
- Stewardship Workflows
- Business Rule Definitions
- Rule Semantics
- **Term to Rule Linking**
- Allowable Values
- Search and Browse
- Data Flow Lineage
- Model Versioning

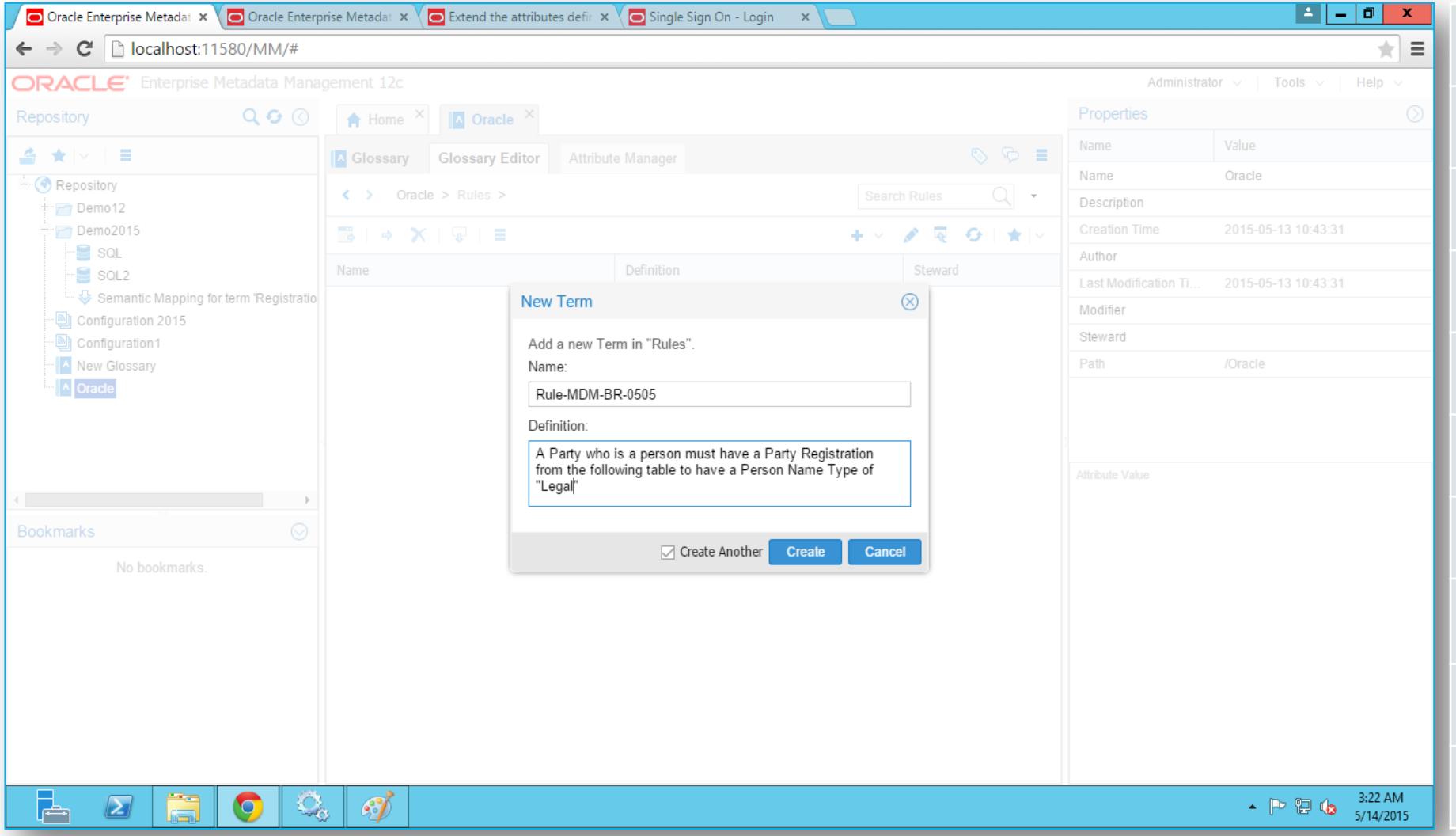


- Model Browser
- Systems Lineage
- Data Profiling
- Data Validation
- Dashboards
- Remediation & Repair
- 3 Key Users
- 3 Kinds of Lineage
- 3rd Party Integrations
- Deployment Architecture

Source : Oracle Product Management and company Information Assets

Capture Business Rule Definitions

- Critical Data Elements
- Business Glossary
- Stewardship Workflows
- **Business Rule Definitions**
- Rule Semantics
- Term to Rule Linking
- Allowable Values
- Search and Browse
- Data Flow Lineage
- Model Versioning

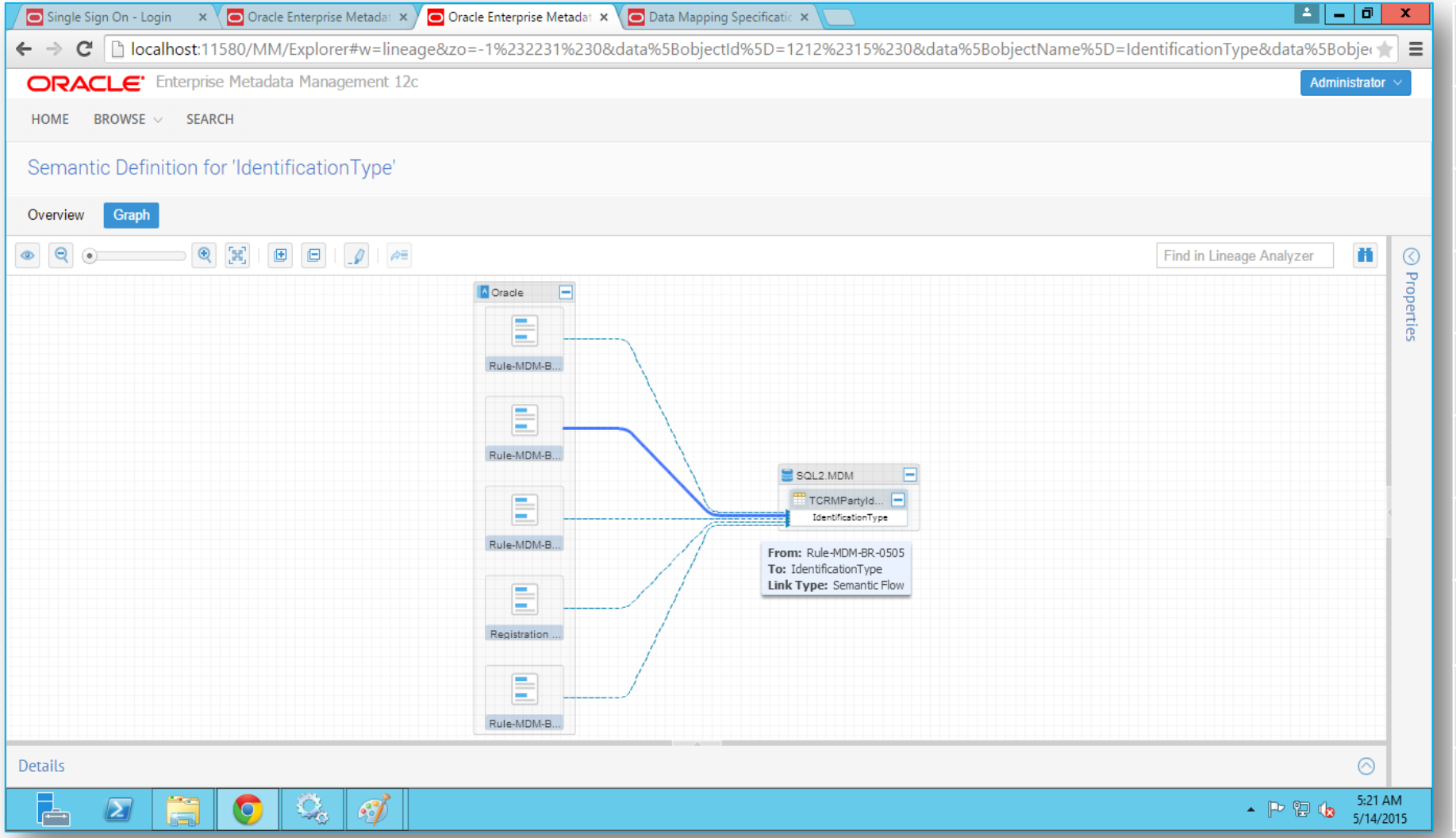


- Model Browser
- Systems Lineage
- Data Profiling
- Data Validation
- Dashboards
- Remediation & Repair
- 3 Key Users
- 3 Kinds of Lineage
- 3rd Party Integrations
- Deployment Architecture

Source : Oracle Product Management and company Information Assets

Semantic Linking of Business Rules

- Critical Data Elements
- Business Glossary
- Stewardship Workflows
- Business Rule Definitions
- **Rule Semantics**
- Term to Rule Linking
- Allowable Values
- Search and Browse
- Data Flow Lineage
- Model Versioning

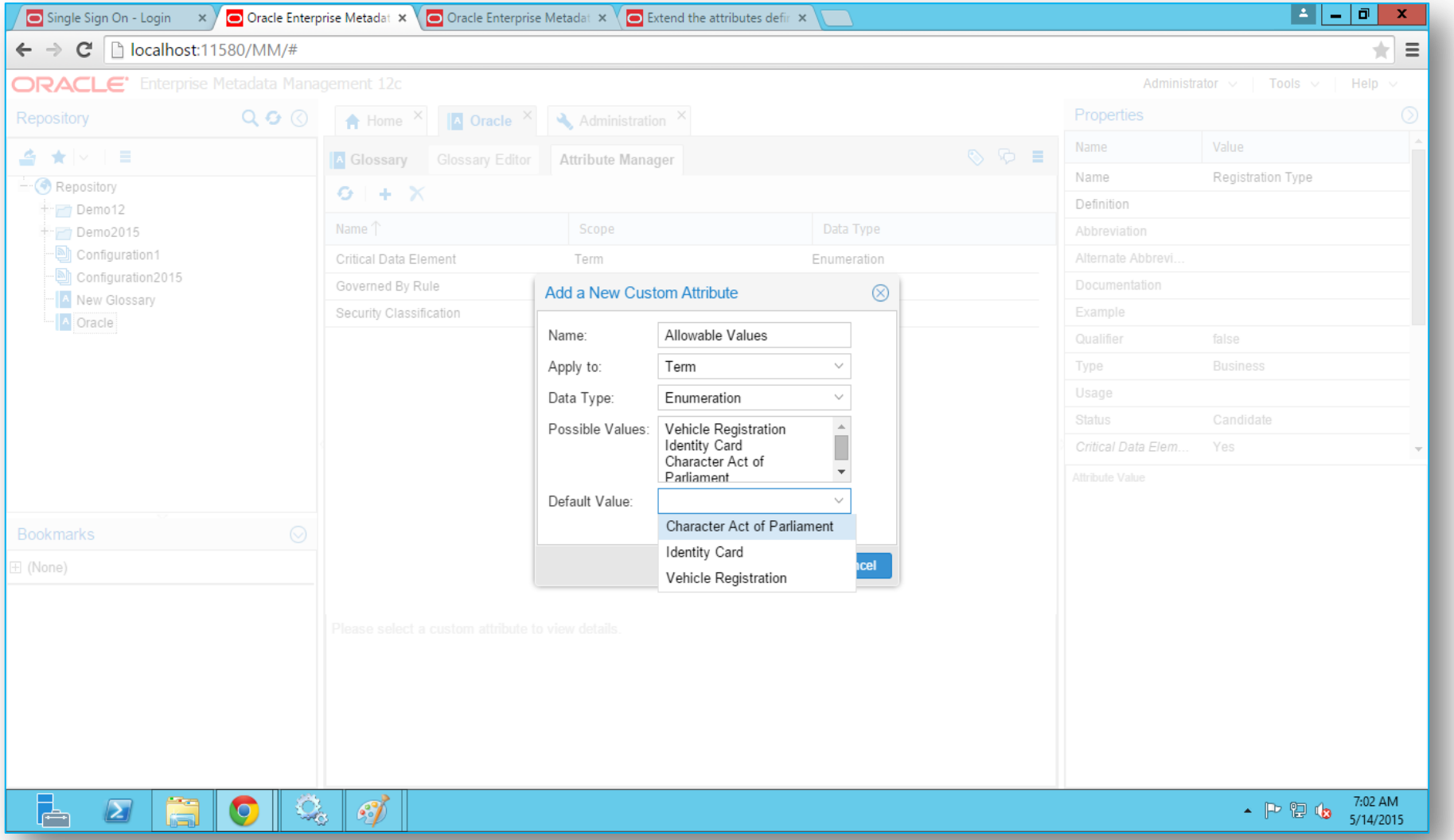


- Model Browser
- Systems Lineage
- Data Profiling
- Data Validation
- Dashboards
- Remediation & Repair
- 3 Key Users
- 3 Kinds of Lineage
- 3rd Party Integrations
- Deployment Architecture

Source : Oracle Product Management and company Information Assets

Allowable Values of Business Terms

- Critical Data Elements
- Business Glossary
- Stewardship Workflows
- Business Rule Definitions
- Rule Semantics
- Term to Rule Linking
- **Allowable Values**
- Search and Browse
- Data Flow Lineage
- Model Versioning

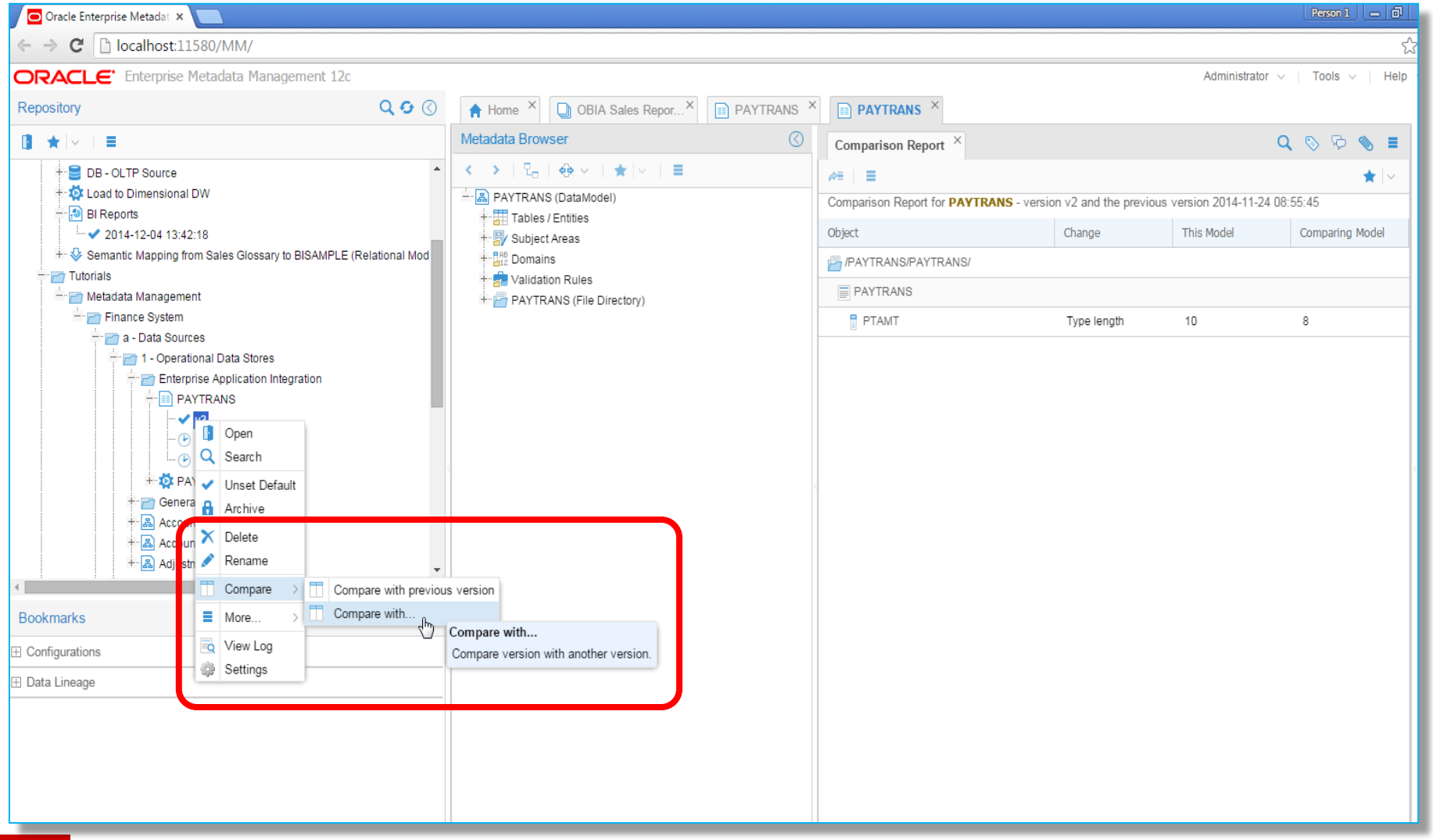


- Model Browser
- Systems Lineage
- Data Profiling
- Data Validation
- Dashboards
- Remediation & Repair
- 3 Key Users
- 3 Kinds of Lineage
- 3rd Party Integrations
- Deployment Architecture

Source : Oracle Product Management and company Information Assets

Model Level Versioning Shows Object Level Changes

- Critical Data Elements
- Business Glossary
- Stewardship Workflows
- Business Rule Definitions
- Rule Semantics
- Term to Rule Linking
- Allowable Values
- Search and Browse
- Data Flow Lineage
- **Model Versioning**

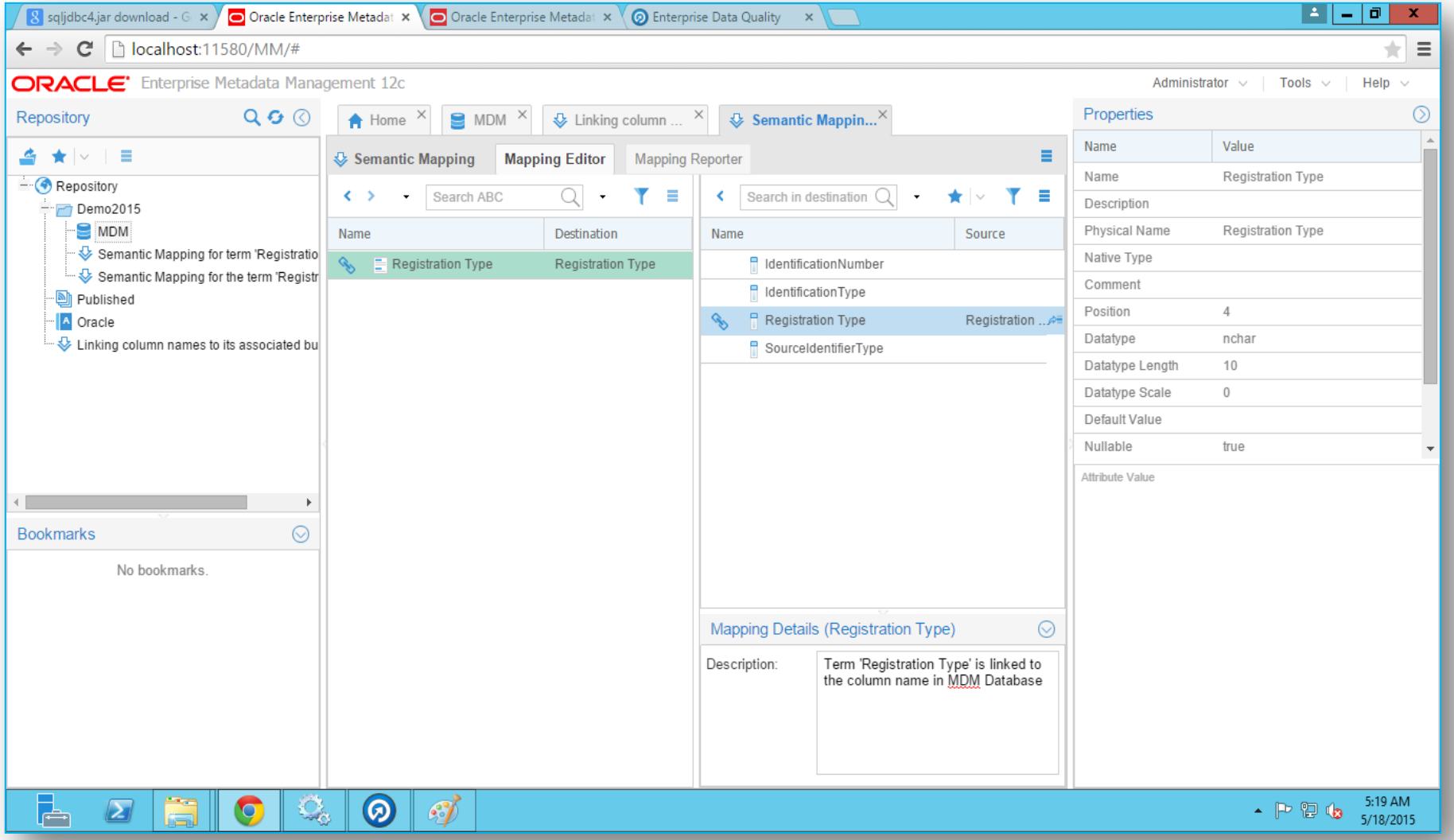


- Model Browser
- Systems Lineage
- Data Profiling
- Data Validation
- Dashboards
- Remediation & Repair
- 3 Key Users
- 3 Kinds of Lineage
- 3rd Party Integrations
- Deployment Architecture

Source : Oracle Product Management and company Information Assets

Vertical Lineage = Business Term to Technical Column

- Critical Data Elements
- Business Glossary
- Stewardship Workflows
- Business Rule Definitions
- Rule Semantics
- Term to Rule Linking
- Allowable Values
- Search and Browse
- Data Flow Lineage
- Model Versioning

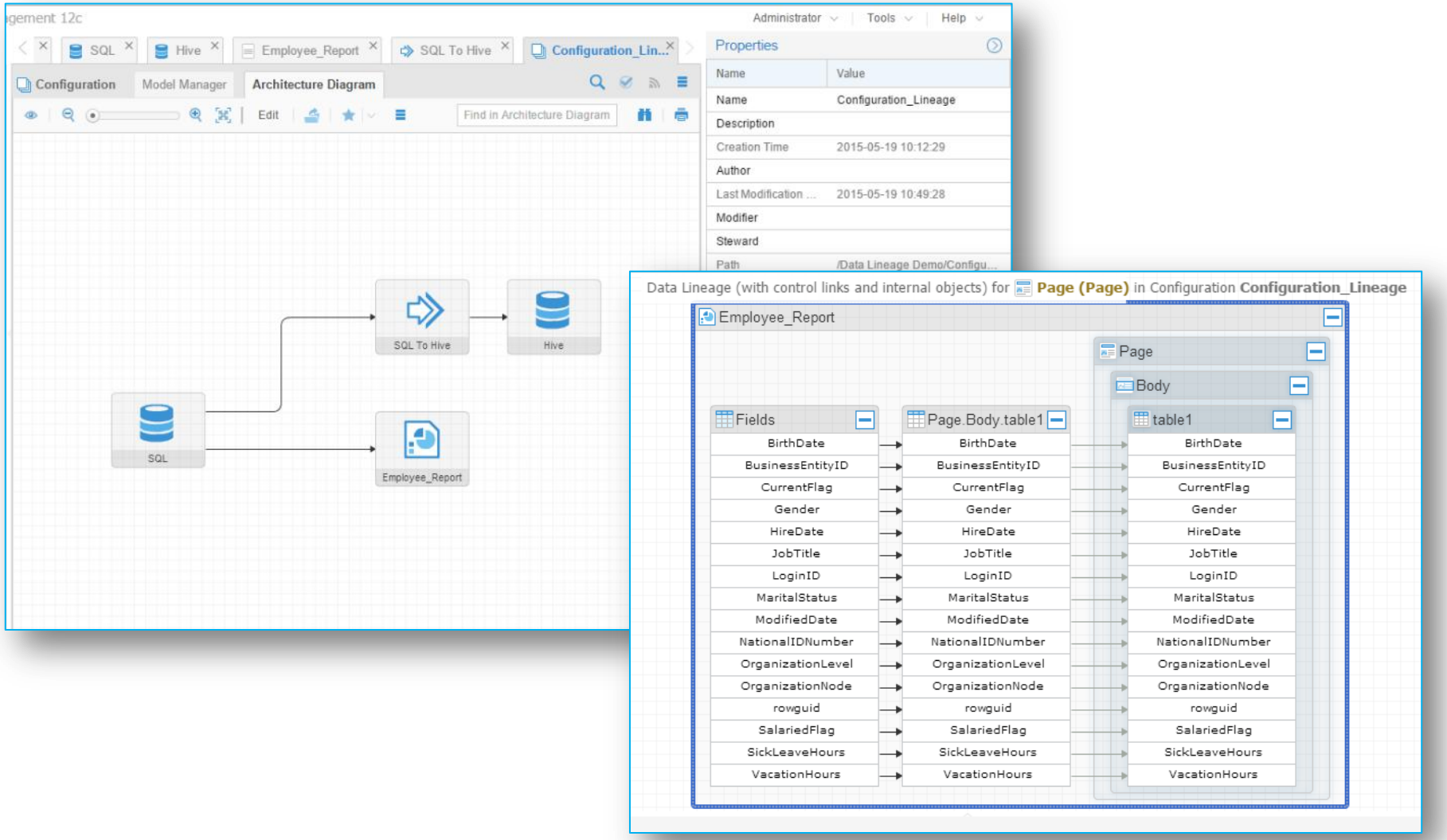


- Model Browser
- **Business & Systems Lineage**
- Data Profiling
- Data Validation
- Dashboards
- Remediation & Repair
- 3 Key Users
- 3 Kinds of Lineage
- 3rd Party Integrations
- Deployment Architecture

Source : Oracle Product Management and company Information Assets

Horizontal Lineage = Report to Source Column / Field Level

- Critical Data Elements
- Business Glossary
- Stewardship Workflows
- Business Rule Definitions
- Rule Semantics
- Term to Rule Linking
- Allowable Values
- Search and Browse
- Data Flow Lineage
- Model Versioning

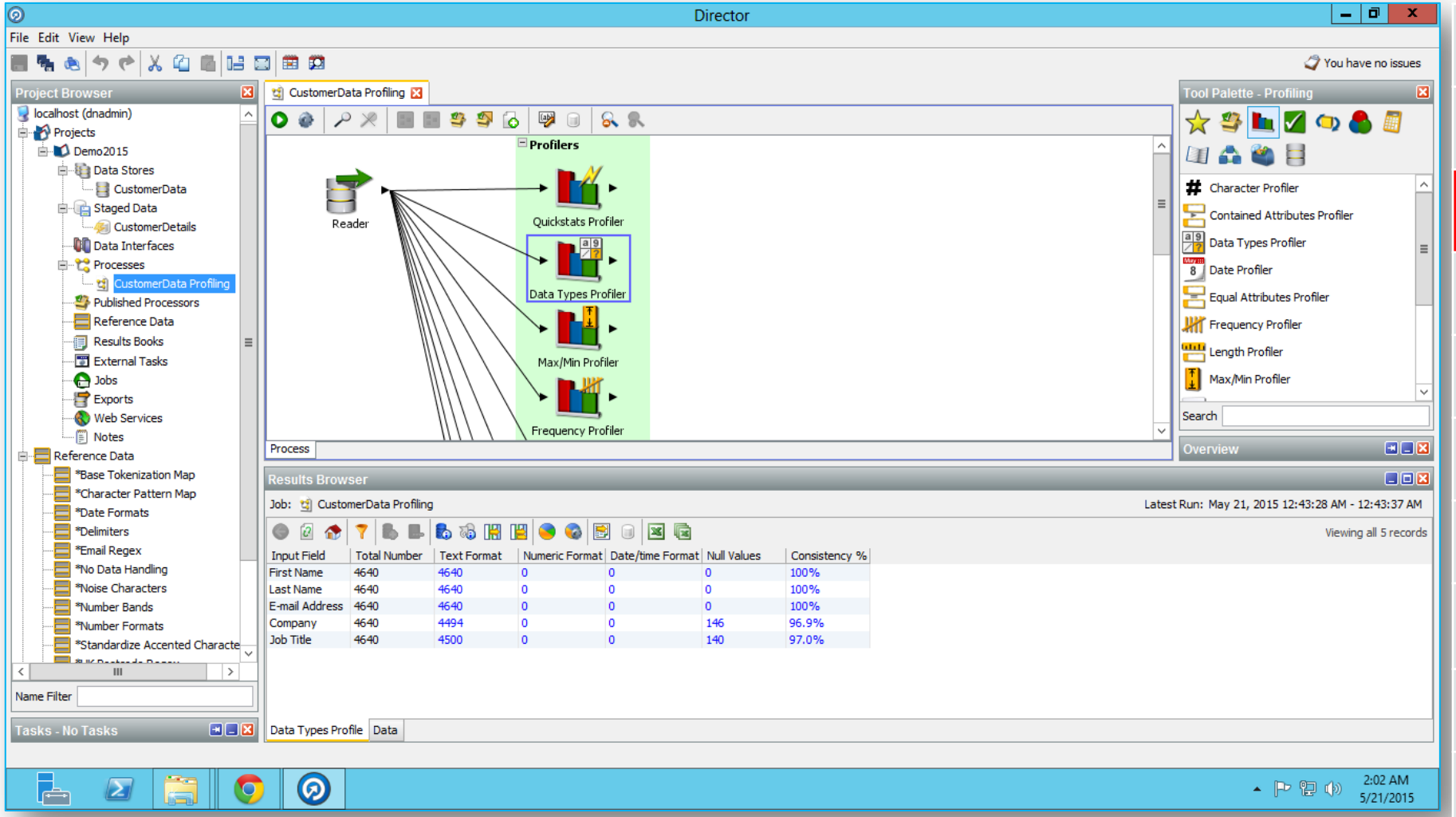


- Model Browser
- **Business & Systems Lineage**
- Data Profiling
- Data Validation
- Dashboards
- Remediation & Repair
- 3 Key Users
- 3 Kinds of Lineage
- 3rd Party Integrations
- Deployment Architecture

Source : Oracle Product Management and company Information Assets

Data Profiling

- Critical Data Elements
- Business Glossary
- Stewardship Workflows
- Business Rule Definitions
- Rule Semantics
- Term to Rule Linking
- Allowable Values
- Search and Browse
- Data Flow Lineage
- Model Versioning

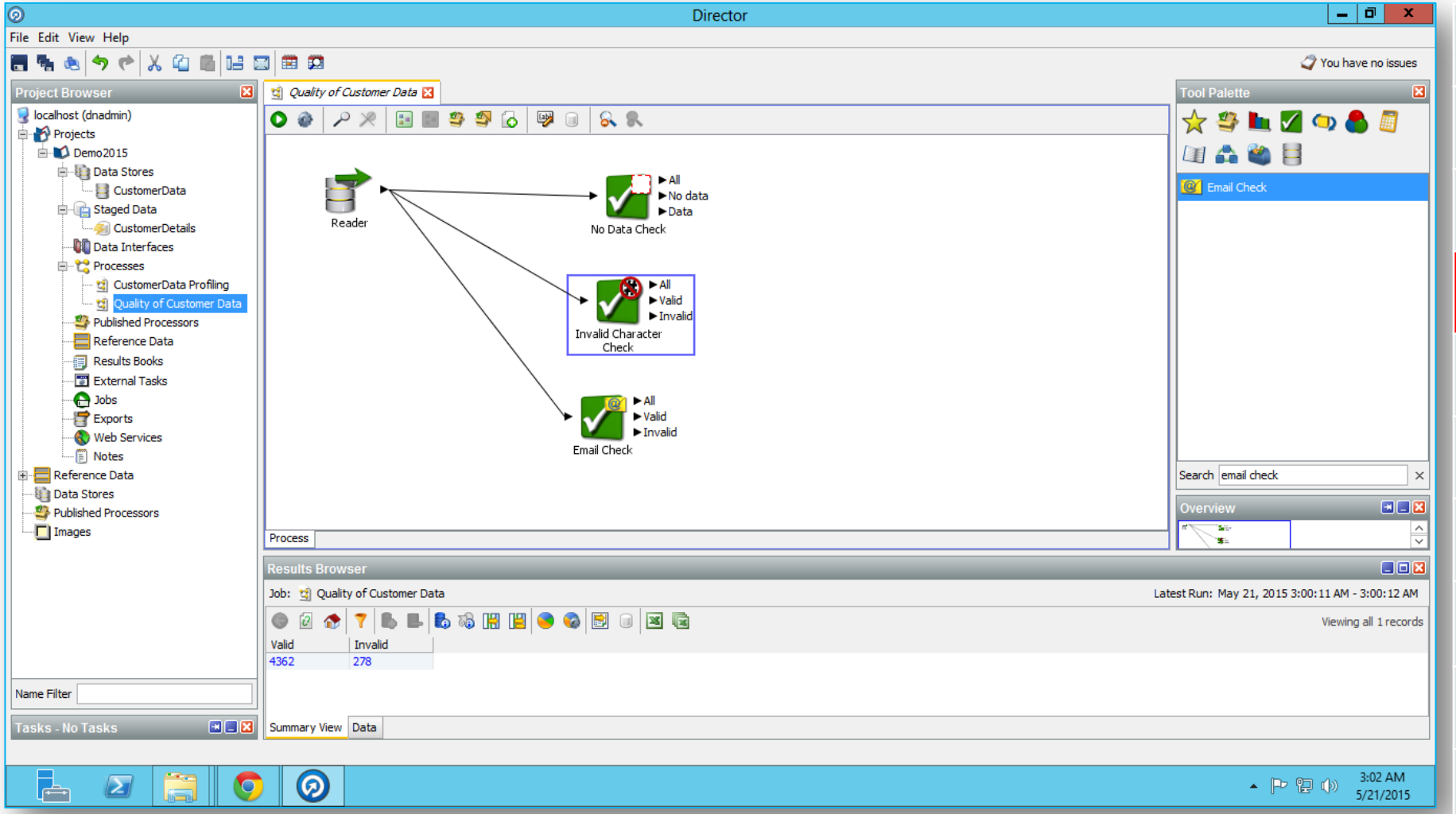


- Model Browser
- Systems Lineage
- **Data Profiling**
- Data Validation
- Dashboards
- Remediation & Repair
- 3 Key Users
- 3 Kinds of Lineage
- 3rd Party Integrations
- Deployment Architecture

Source : Oracle Product Management and company Information Assets

Data Validation Against Business Rules

- Critical Data Elements
- Business Glossary
- Stewardship Workflows
- Business Rule Definitions
- Rule Semantics
- Term to Rule Linking
- Allowable Values
- Search and Browse
- Data Flow Lineage
- Model Versioning

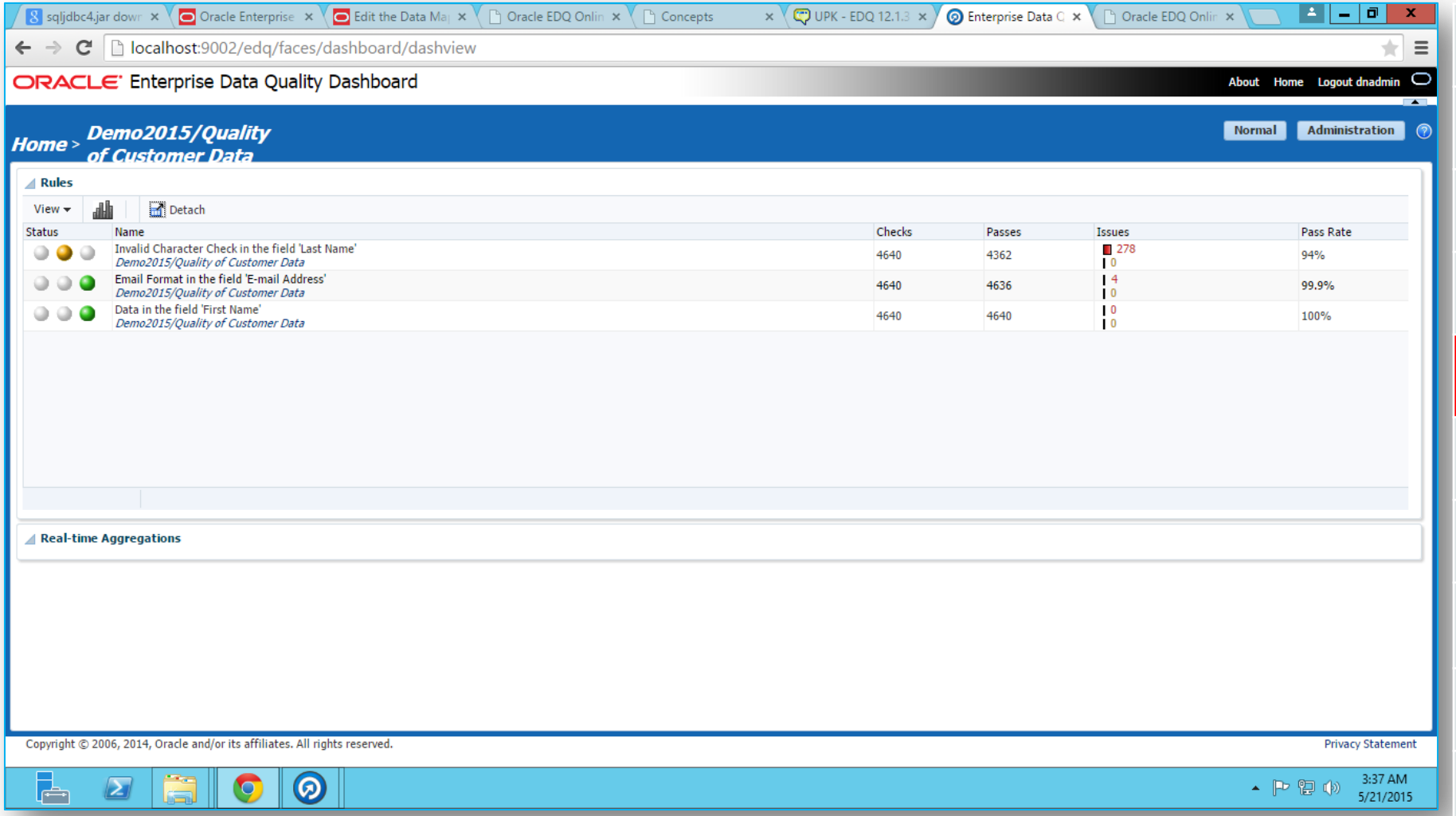


- Model Browser
- Systems Lineage
- Data Profiling
- **Data Validation**
- Dashboards
- Remediation & Repair
- 3 Key Users
- 3 Kinds of Lineage
- 3rd Party Integrations
- Deployment Architecture

Source : Oracle Product Management and company Information Assets

Data Quality Dashboard

- Critical Data Elements
- Business Glossary
- Stewardship Workflows
- Business Rule Definitions
- Rule Semantics
- Term to Rule Linking
- Allowable Values
- Search and Browse
- Data Flow Lineage
- Model Versioning

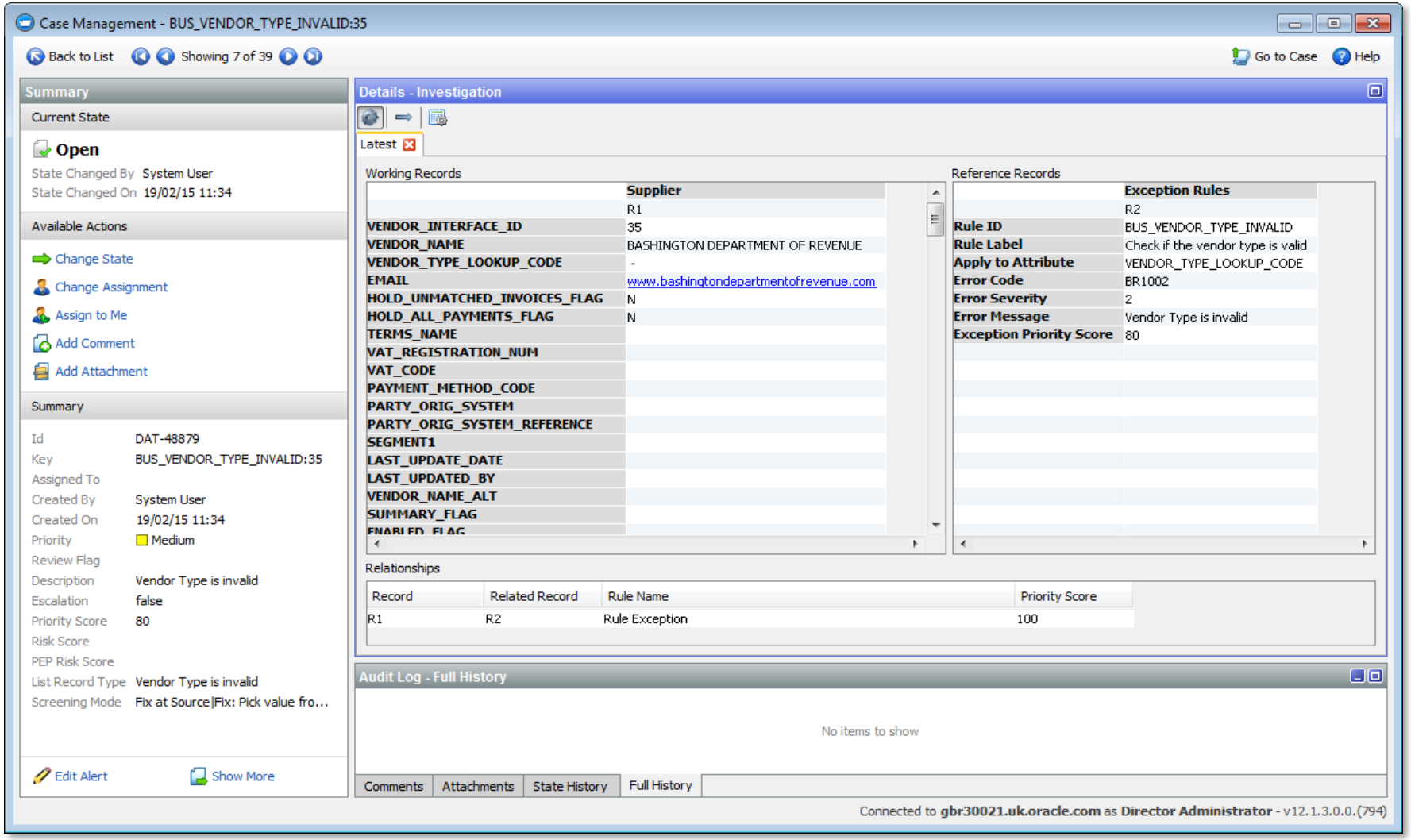


- Model Browser
- Systems Lineage
- Data Profiling
- Data Validation
- **Dashboards**
- Remediation & Repair
- 3 Key Users
- 3 Kinds of Lineage
- 3rd Party Integrations
- Deployment Architecture

Source : Oracle Product Management and company Information Assets

Data Remediation and Repair

- Critical Data Elements
- Business Glossary
- Stewardship Workflows
- Business Rule Definitions
- Rule Semantics
- Term to Rule Linking
- Allowable Values
- Search and Browse
- Data Flow Lineage
- Model Versioning



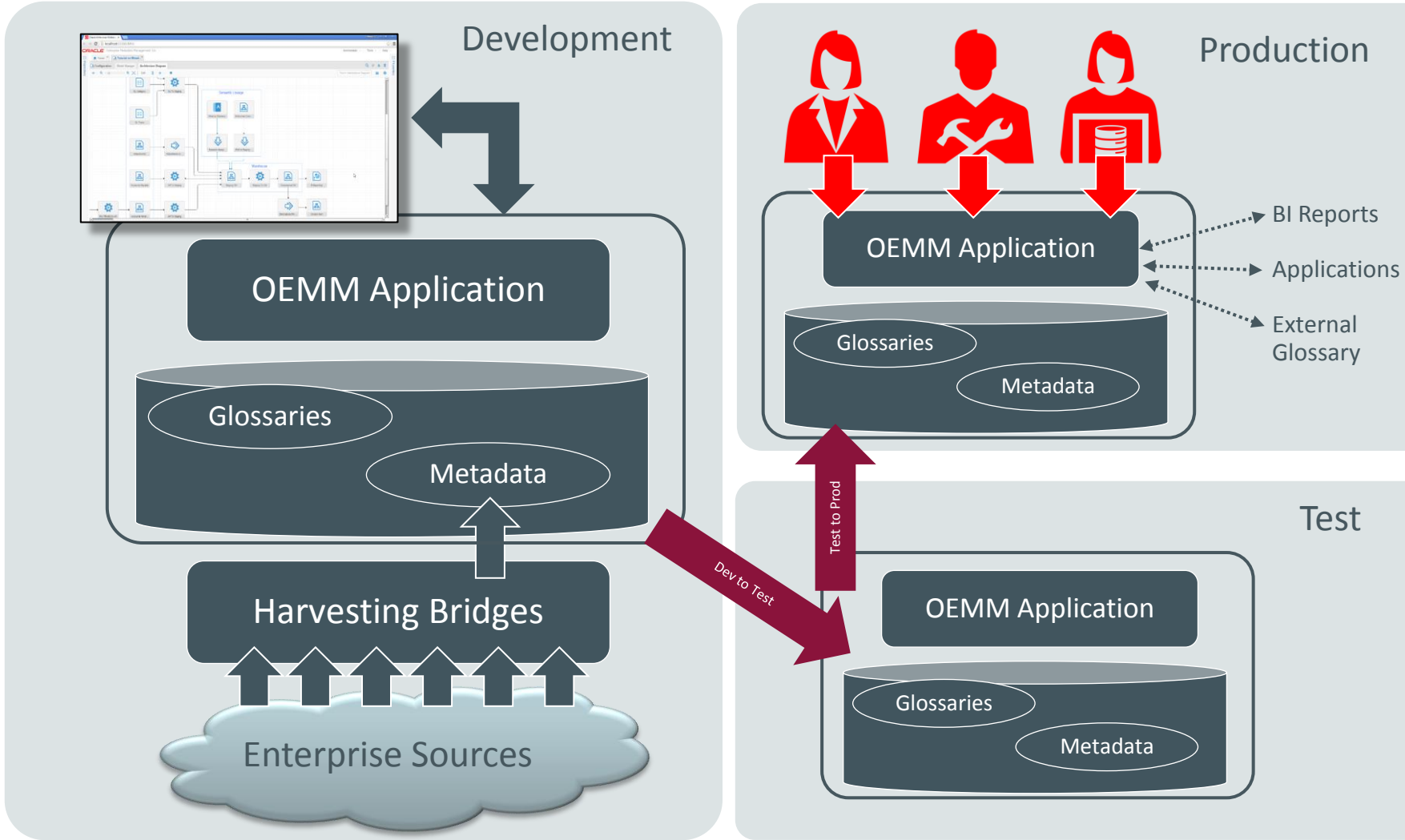
- Model Browser
- Systems Lineage
- Data Profiling
- Data Validation
- Dashboards
- **Remediation & Repair**
- 3 Key Users
- 3 Kinds of Lineage
- 3rd Party Integrations
- Deployment Architecture

ORACLE

Source : Oracle Product Management and company Information Assets

Deployment Architecture

- Critical Data Elements
- Business Glossary
- Stewardship Workflows
- Business Rule Definitions
- Rule Semantics
- Term to Rule Linking
- Allowable Values
- Search and Browse
- Data Flow Lineage
- Model Versioning



- Model Browser
- Systems Lineage
- Data Profiling
- Data Validation
- Dashboards
- Remediation & Repair
- 3 Key Users
- 3 Kinds of Lineage
- 3rd Party Integrations
- **Deployment Architecture**