



DATA GOVERNANCE AND MASTER DATA MANAGEMENT CONFERENCE EUROPE

11 - 14 March 2024 | London, UK

****Please score and comment on this session and speaker
in the event mobile app****



The Value of AI in Master Data Management

Mike Ferguson
Managing Director
Intelligent Business Strategies




IRM EDG and MDM Conference
March 2024

Who Is Mike Ferguson? – A Leading Analyst in Data Management & Analytics

Edgar F Codd

Turing Award Winner




Relational Model

All relational DBMSs SQL Language Data Normalisation

Codd & Date


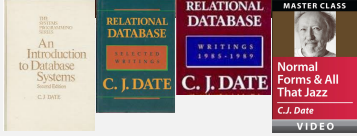
Mike Ferguson




Chief Architect teradata.

1st massively parallel relational DBMS

C J Date

Founded 1992



Copyright © Intelligent Business Strategies 1992-2024

Mike Ferguson Is a Leading Industry Analyst / Consultant and Conference Chairman of Big Data LDN - Leading The Industry In Data Management and Analytics



Big Data LDN is the largest data & analytics conference in Europe

- 20000 delegates
- 180 vendors
- 15 theatres
- 300+ speakers

Copyright © Intelligent Business Strategies 1992-2024



About Intelligent Business Strategies

- UK-based independent IT analyst and consulting firm founded 1992 specialising in data management and analytics
- Mike Ferguson is co-founder of the Data and Analytics Retreat and Conference Chairman Big Data LDN
- Three main lines of business

Research

- Market research
 - 4th Industrial Revolution Survey
- D&A product research
 - Data Catalogs
 - Data Fabric
 - Data Governance
 - Master Data Management
 - Distributed DBMS
 - Analytical DBMS

Education

- Data Strategy
- How to Govern Data Across a Distributed Data Landscape
- Practical Guidelines for Implementing a Data Mesh
- Modern Data Architecture
- DW Modernisation
- DW Migration to the Cloud
- Embedded Analytics, Intelligent Apps & AI Automation
- Public classes (anyone)
- On-site classes (single client)
 - Customers, vendors, systems integrators
- On-line (public & on-sites)

Consulting

- Customers
 - D&A Strategy, Data Architecture
 - D&A Technology selection
 - D&A Reviews, Data Governance
 - Implementation Project advisory
- Vendors
 - Product strategy
 - Product positioning
 - Marketing support
 - Speaking at vendor events
 - White papers
 - Webinars
- Venture Capitalists
 - Due-diligence, Asset advisory

www.intelligentbusiness.biz

Copyright © Intelligent Business Strategies 1992-2024

5

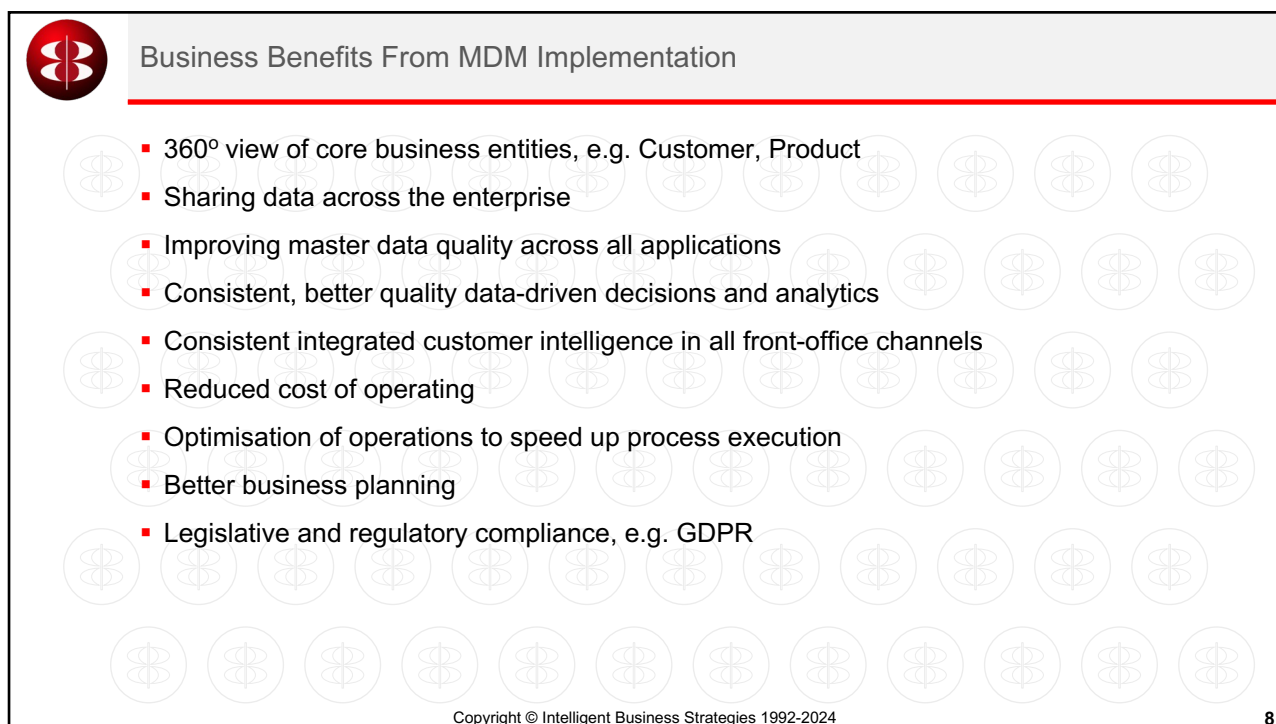
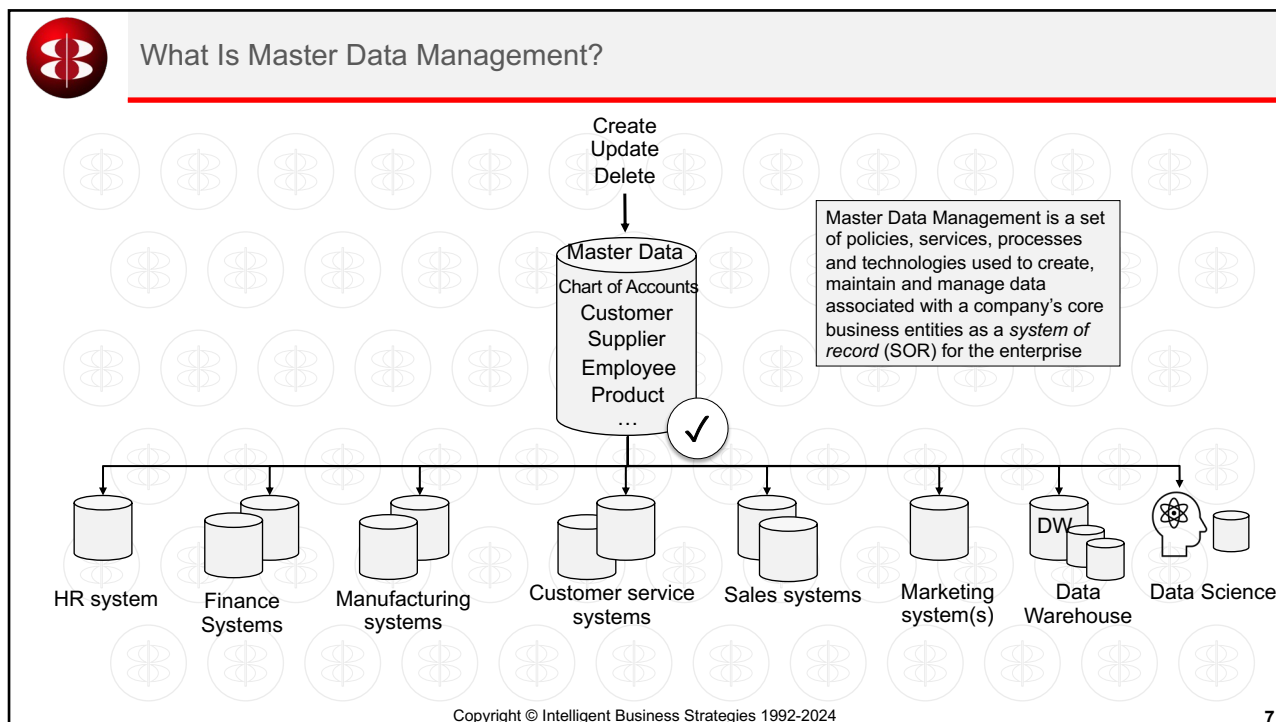


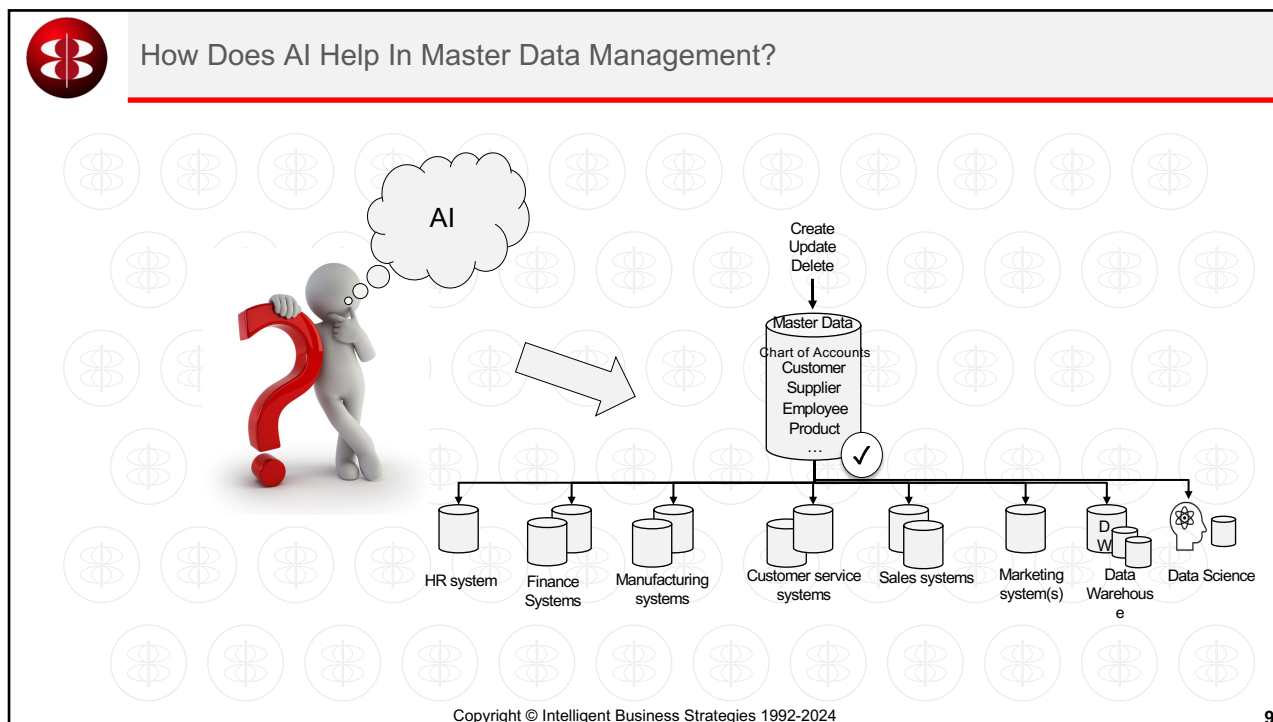
Topics

- What is master data management?
- How does AI help in building master data entities?
 - Automated enrichment of master data business terms in a business glossary
 - Using generative AI to build a data models for master data
 - Using AI to auto discover physical master data and map it to a business glossary
 - Automating sensitive data detection and master data metadata enrichment using AI
 - Using generative AI to create data quality rules to detect poor quality source data
 - AI-assisted data cleansing and join recommendations in data engineering
 - Using AI to automate matching to create master data
 - Using Generative AI to build pipelines to create master data products
- Generative AI to automatically create product information and marketing content from master data

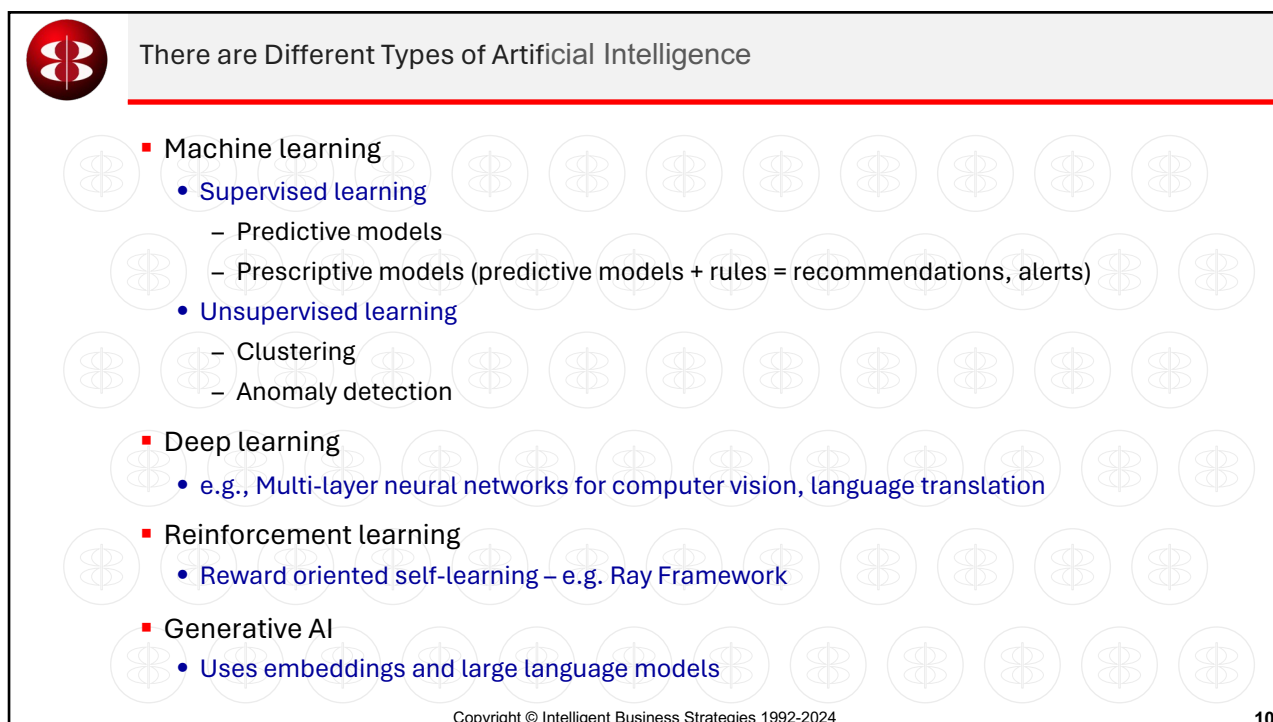
Copyright © Intelligent Business Strategies 1992-2024

6





9



10



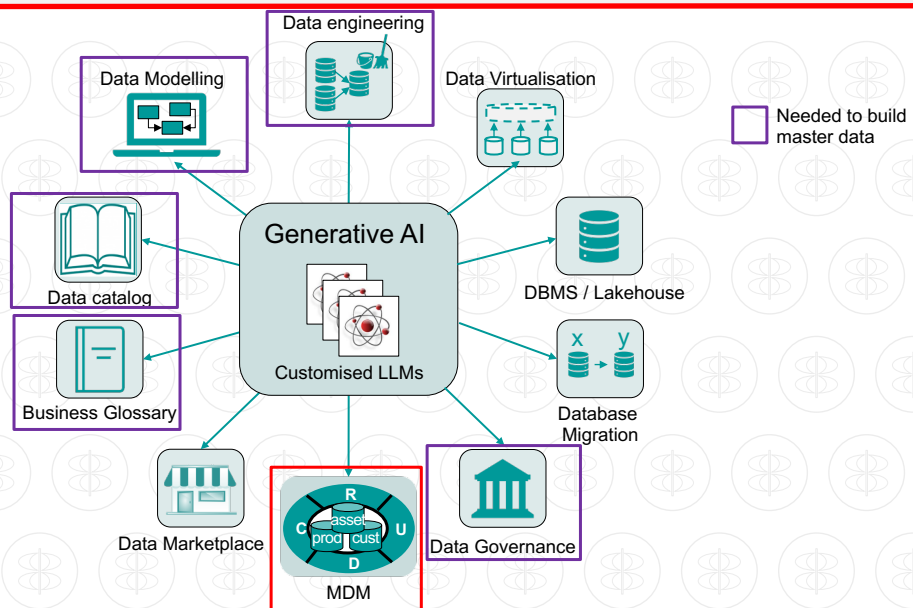
What is Generative AI?

What is Generative AI?
 A subset of deep learning where multi-layer neural network models **generate new content** such as text, images, audio, video, code, and synthetic data in response to prompts based on what the models have learned from patterns in the content they were trained on

- Foundation Large Language Model examples:
 - AWS Titan, Databricks Dolly, Google PaLM2 & Codex,
 - Open AI Generative Pre-Trained Transformer (GPT4)
 - Anthropic Claude
- General use cases
 - Text generation
 - Virtual assistant – chat
 - Conversational search
 - Summarisation – text extraction
 - Code generation
 - Synthetic data generation
 - Image generation and classification
 - Video generation
- Benefits
 - Improve customer and employee experience
 - Productivity, automation
 - Ease of use, lower skills bar
 - Democratisation of D&A development



Generative AI Has Emerged In Almost Every Aspect of Data Management



A Methodology Is Needed To Build Master Data Entities When Implementing MDM - AI Can Help Improve Speed of Implementation

Customer Data Governance

Product Data Governance

Supplier Data Governance

SBV = Shared Business Vocabulary

Also ASSET
EMPLOYEE
ACCOUNT
MATERIAL
.....

What steps in the methodology can be improved by AI?

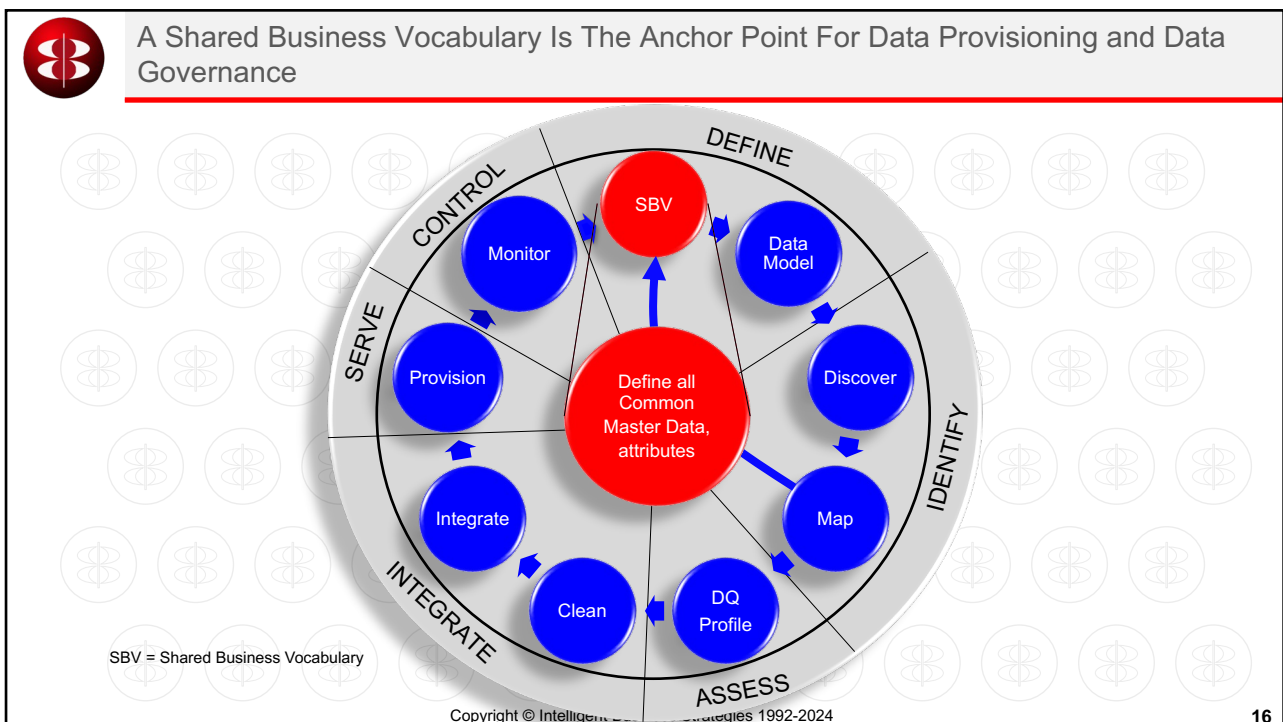
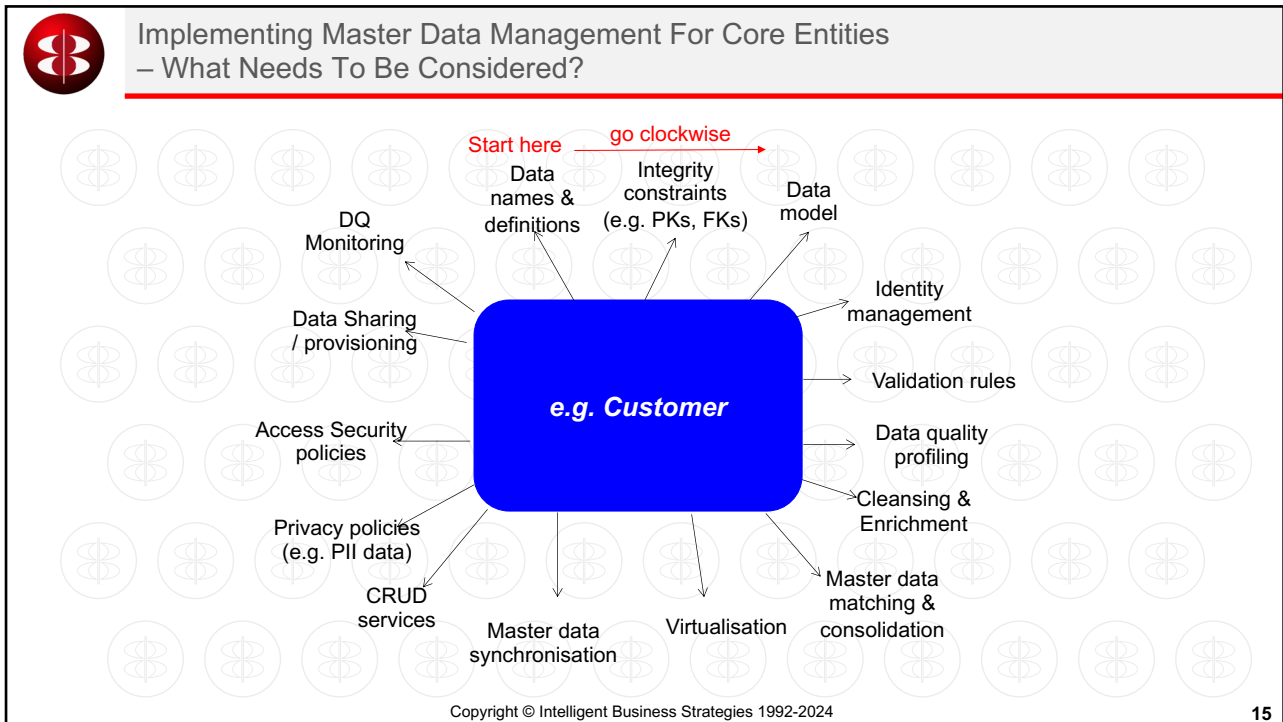
Copyright © Intelligent Business Strategies 1992-2024

Master Data Governance Policies Also Need To Be Defined

- Several types of policy need to be defined to govern master data
 - Master data integrity constraints
 - Master data quality validation policies
 - Master data cleansing rules
 - Master data integration rules
 - Master data provisioning policies, e.g. Language rendering, formats, etc.
 - Master data access control policies (CRUD security)
 - Master data privacy policies
 - Master data sharing policies
 - Master data lifecycle (expiration) policies
- These policies can be applied to:
 - Individual data items
 - All data items in a master data entity, e.g. Customer, Asset, Product
 - Both
- Policies may be restricted to limit their governance capability
 - e.g. limited to master data in a specific application or process

How does AI help in governing master data?

Copyright © Intelligent Business Strategies 1992-2024

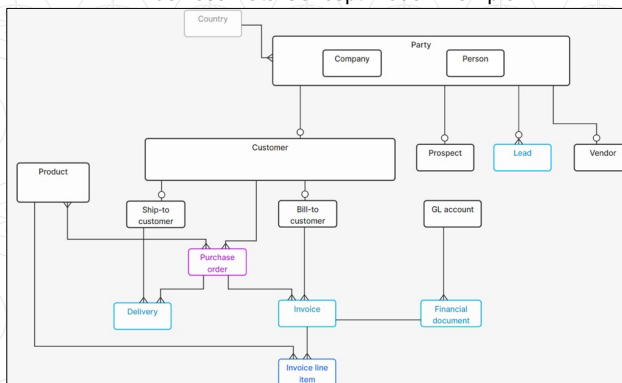




A Good Way To Start A Business Glossary Is To Create A Business Data Concept Model

- Identify the data concepts, properties and relationships and construct a data concept model
- It is good practice to highlight all master and transaction data concepts in your data concept model

Business Data Concept Model Example



Some vendors provide complete definitions for business data entities and all their attributes to get you started quickly, e.g., Collibra uses Schema.org, Microsoft has a CDM

Copyright © Intelligent Business Strategies 1992-2024

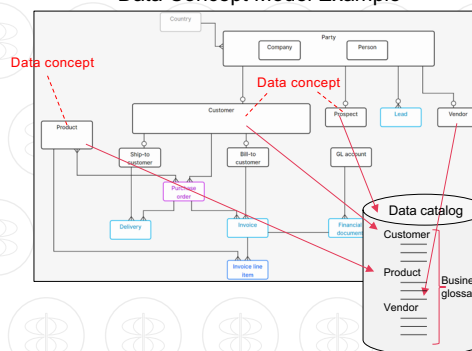
17



Steps To Creating A Common Vocabulary – From Data Concept Model To Master Data Product

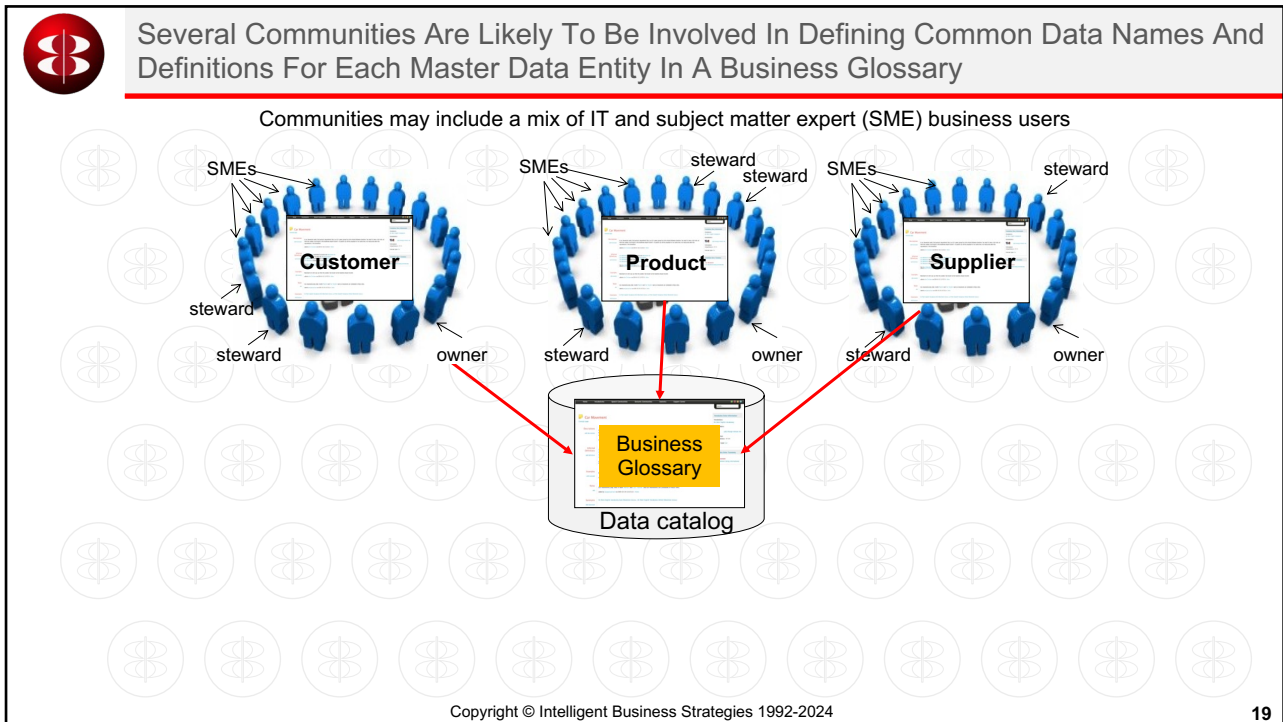
1. Identify the data concepts, properties and relationships and construct a data concept model
2. The data concepts become the 'skeleton data entities' in the common vocabulary
3. Each data concept and its attributes should be defined in the business glossary as a data entity with a data owner
 - Including master data entities
4. Use the catalog to discover the data for each master data entity in underlying data stores across the data landscape
5. Design DataOps component-based pipelines to create the master data entities (master data products) with common vocabulary data names and store them in an MDM system
6. Publish all master data entities stored in the MDM system as data products in a data marketplace

Data Concept Model Example



Copyright © Intelligent Business Strategies 1992-2024

18



Using Generative AI For Data Catalog Metadata Enrichment – Auto-Generation Of Business Term Descriptions In A Business Glossary In Atlan

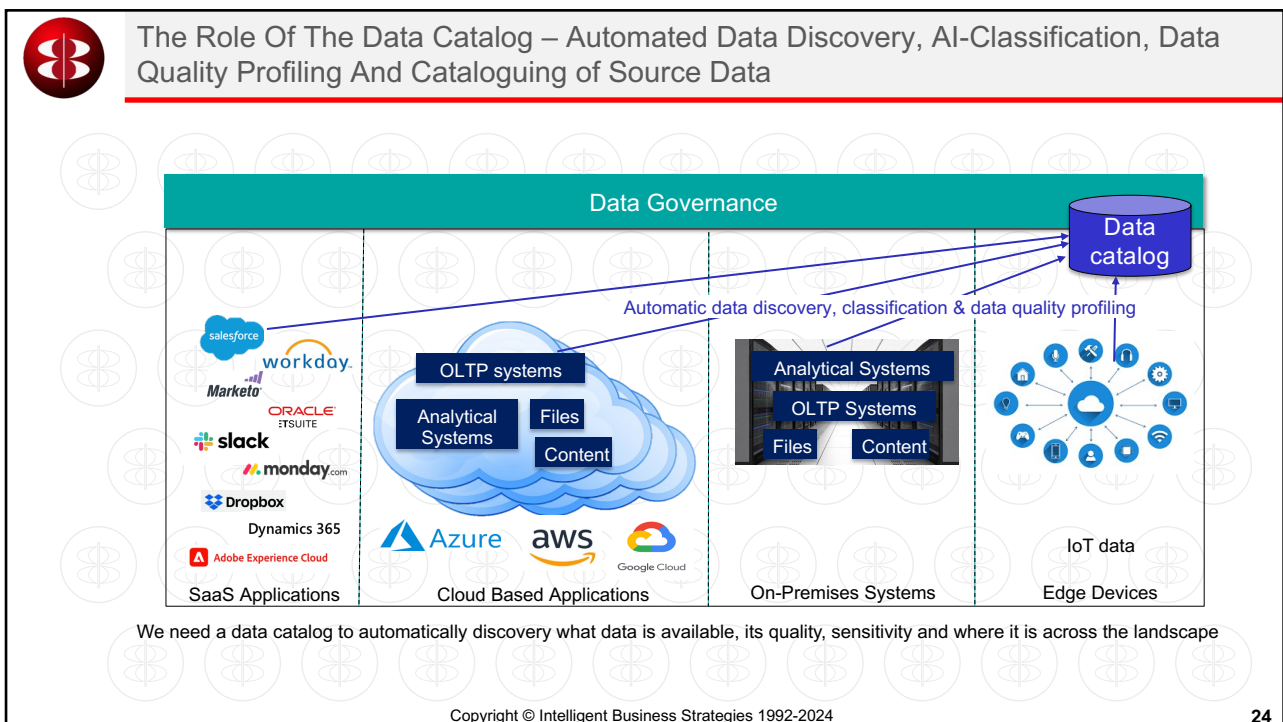
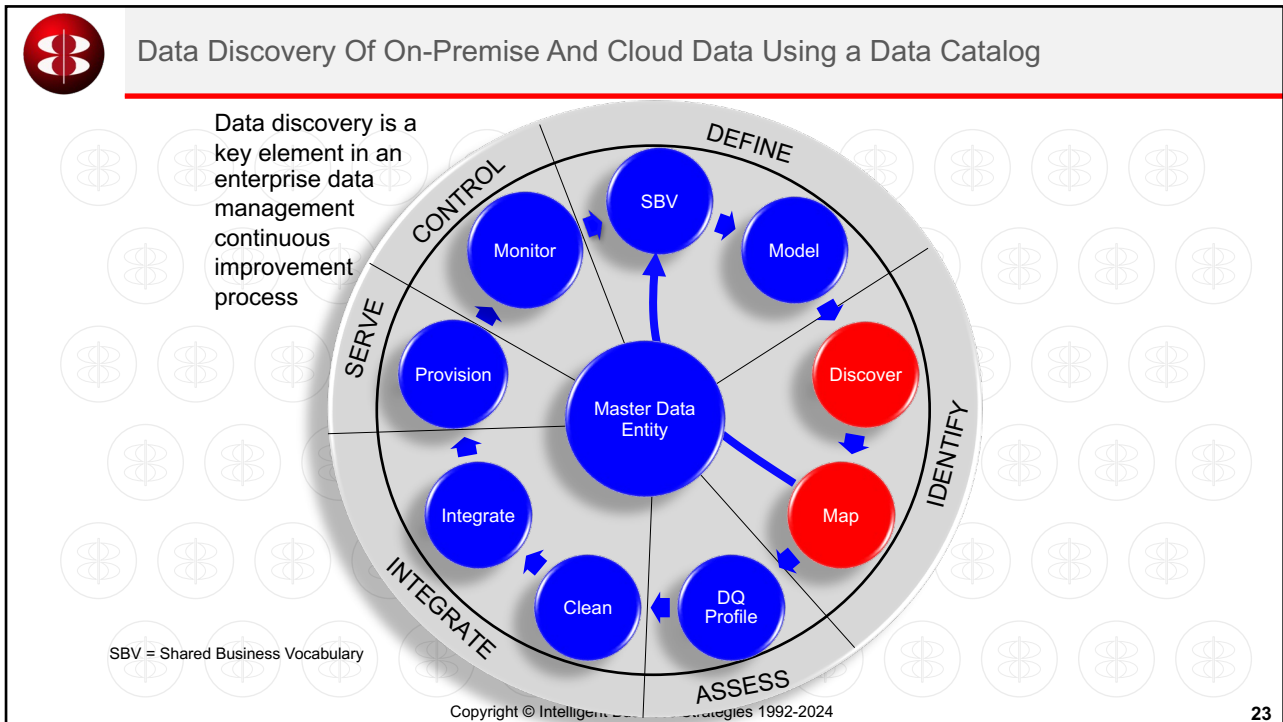
The screenshot shows the Atlan Business Glossary interface. On the left, a 'Term Summary' for 'Claim approval rate' is visible. A red box highlights the 'Add README' button (1). A modal window titled 'Pick a template' is open, showing 'Glossary Term README' as a selected option (2). A red box highlights the 'Use Atlan AI' button (3). An arrow labeled 'AI generate' points from the modal to the right-hand side of the screenshot, which shows the generated README text for 'Claim approval rate'.

Business Glossary Data catalog

Source: Atlan

Add a Readme, pick from a template and auto generate the README text using a generative LLM

Copyright © Intelligent Business Strategies 1992-2024 20





Data Catalog Products – A Very Crowded Market

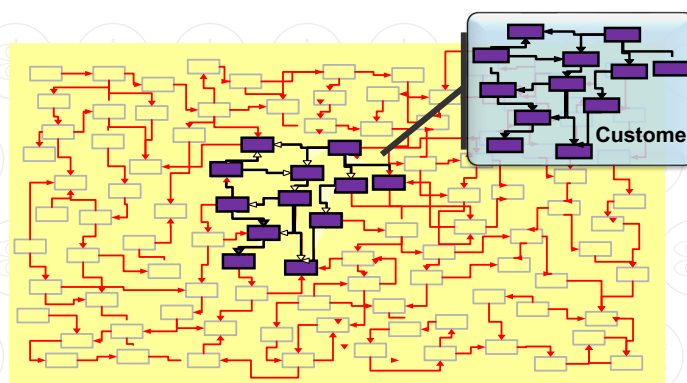
- Alation
- Alex Solutions
- Alteryx Connect
- Amazon Glue Catalog
- Apache Atlas (open source)
- Atlan
- Ataccama ONE Data Catalog
- BigID Data Catalog
- Boomi AtomSphere Data Platform Catalog
- Cambridge Semantic Anzo Catalog
- Cloudera Data Platform SDX Catalog
- Collibra Catalog
- Databricks Unity Catalog
- data.world
- Denodo Catalog
- Google Cloud Data Catalog
- Hitachi Vantara Lumada Data Catalog
- IBM Watson Knowledge Catalog
- Informatica Enterprise Data Catalog
- Microsoft Purview
- Oracle Cloud Infrastructure Data Catalog
- Qlik
 - Enterprise Data Catalog
 - Talend Data Catalog
- Quest (formerly erwin) Data Catalog
- Rocket Software (formerly ASG) Intelligent Data Catalog
- SAP Datasphere Catalog (cloud)
- SAS Information Catalog
- Salesforce Tableau Catalog
- TIBCO Cloud Metadata Catalog
- Top Quadrant TopBraid EDG Data Catalog
- Truist Zoloni Arena Data Catalog

Copyright © Intelligent Business Strategies 1992-2024

25



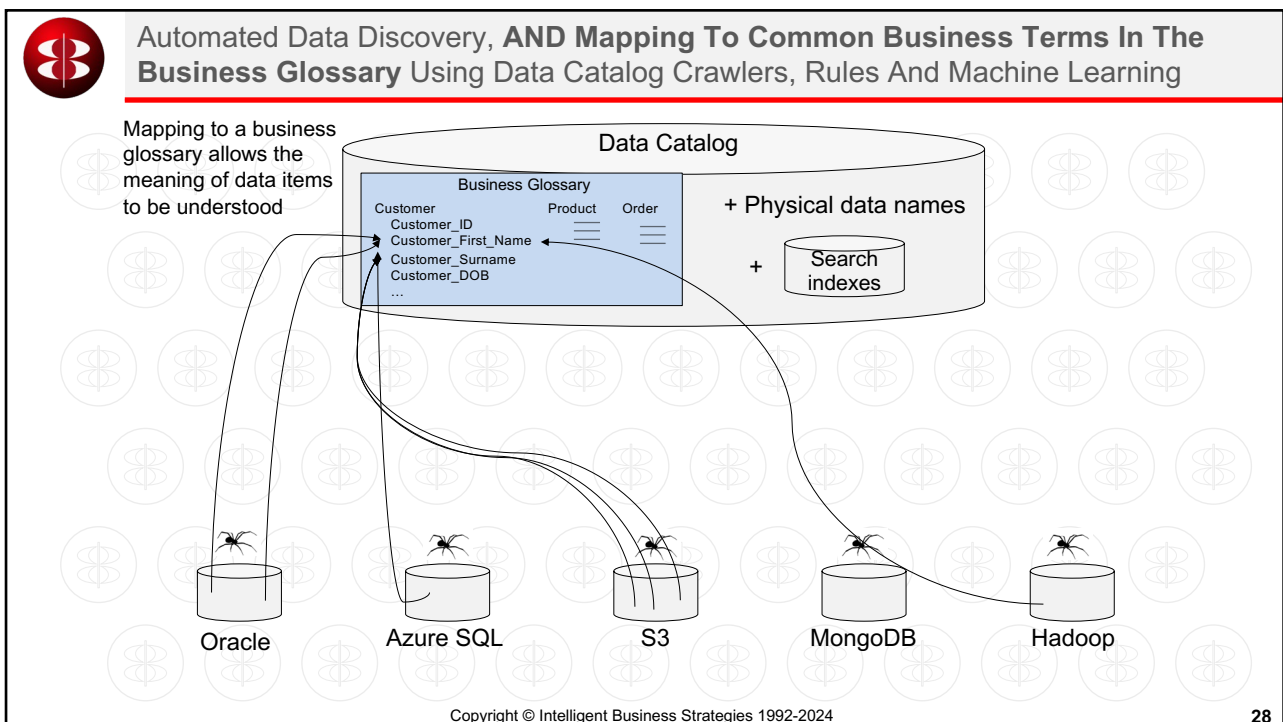
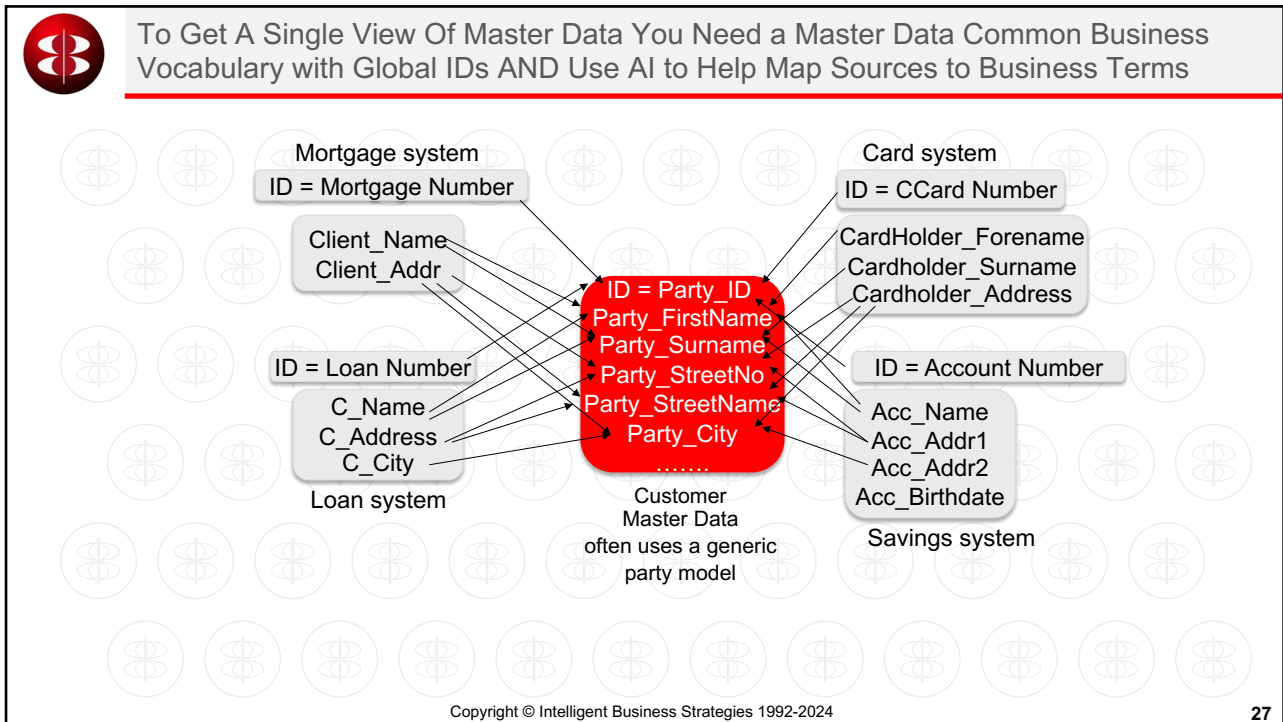
Human Assisted AI Automated Data Discovery Seeks To Determine What Data You Have And Find Complete Data Objects Across Heterogeneous Systems



- Automated grouping of tables across multiple systems into business entities

Copyright © Intelligent Business Strategies 1992-2024

26



Enabling Glossary Business Term Association In Informatica IDMC Data Governance and Catalog To Map Physical Data Names To Business Terms in a Business Glossary

Informatica Metadata Command Center

Amazon Redshift

1 Registration 2 Configuration 3 Associations 4 Schedule

Metadata Extraction Data Profiling and Quality Data Classification Relationship Discovery **Glossary Association**

Enable Glossary Association:

Enable auto-acceptance Confidence Score Threshold for Auto-Acceptance: 90 %

Considers all published business terms in the glossary while making recommendations to associate your technical assets.

Informatica, can automatically map physical data names to business terms in a business glossary. It does this by using an algorithm (based on accepted business terms on data domains, column similarity, and name match between a column and business term) to calculate a confidence score that it can map it correctly. You can specify a threshold for the confidence score above which it will automatically assign a business term to a physical data asset.

Source: Informatica

Copyright © Intelligent Business Strategies 1992-2024

29

Informatica MDM AI-Driven Smart Field Mapping – Based on Name Similarity

- Exact Field Name mapping - Matches fields of the same name.
- Smart Field Mapping - Matches fields with similar names
 - E.g. If an incoming field `Cust_Name` and a target field `Customer_Name`, it automatically links the `Cust_Name` field with the `Customer_Name` field

Informatica Customer 360

File Import

1 Upload File 2 Map Fields 3 Preview and Import File

Map the columns from your source file to suitable target business entity fields, relationships and Hierarchies. If you use any field in a field group, ensure that you map all the required fields of the field group.

Source (Sites.csv)

Source Columns	Row 1 of first 5	Map
SOURCEKEY	103071255	Map
Name	Store #103071255	Map
First Opened	2013-08-16	Map
Sq Footage	1000	Map
Curbside Pickup	TRUE	Map
Postal Address Primary Key	103071255	Map
ADDRESSLINE1	800 S Buena Vista Dr	Map
ADDRESSLINE2	Mailing	Map
CITY	Burbank	Map
COUNTRY	US	Map
DEFAULTINDICATOR	TRUE	Map
POSTALCODE	91521	Map
POSTALCODEEXTENSION		Map
STATE	CA	Map

Target

Target Field	Source Column Header	Map
Source Primary Key*	SOURCEKEY	X
Name*	Name	X
Date Open		
Sq Footage	Sq Footage	X
Curbside Pickup	Curbside Pickup	X
Address		
Address Source Primary Key*	Locations > Address	
Default Indicator	DEFAULTINDICATOR	X
Address Type	ADDRESSSTYPE	X
Usage Type	USAGETYPE	X
Address Line 1*	ADDRESSLINE1	X
Address Line 2		
Address Line 3		

Source: Informatica

Copyright © Intelligent Business Strategies 1992-2024

30



Generative AI In Data Catalogs – Automated Metadata Enrichment Using Atlan AI To Generate A Description Of An Asset

The screenshot shows the Atlan interface for the FCT_ORDERS table. The 'Description' field is empty, and a button labeled 'Request Atlan AI to generate one' is visible. A table of sample data is shown below.

#	ORDER_ID	ORDER_LINE_ID	ORDER_DESCRIPTOR	STOCK_ITEM_ID	STO
1	45	1	32 mm Double sided bu...	164	32 mm
2	1	2	Ride on toy sedan car (BL...	67	Ride on
3	2	3	Developer joke mug - ol...	50	Develop
4	46	4	"The Gu" red shirt XML t...	89	"The Gu
5	46	5	32 mm Anti static bubbl...	171	32 mm
6	2	6	USB food flash drive - ch...	10	USB for
7	47	7	10 mm Anti static bubbl...	167	10 mm
8	47	8	Void fill 400 L bag (Whit...	219	Void fill
9	3	9	Superhero action jacket ...	114	Superh

FCT_ORDERS table currently has no description

Request Atlan AI to generate one



Source: Atlan

Copyright © Intelligent Business Strategies 1992-2024

31



Atlan Generated Contextual Description – E.g. The Description Includes The Database The Table Is In, The Schema

The screenshot shows the Atlan interface for the FCT_ORDERS table. The 'Description' field now contains an AI-generated contextual description. A button labeled 'Apply' is visible, and a note states 'AI generated content may be incorrect'.

AI Suggested

The FCT_ORDERS table in the ANALYTICS database and WIDE_WORLD_IMPORTERS schema contains information about orders placed by customers. It includes details such as the brand, picked quantity, unit price, order date, expected delivery date, customer name, customer purchase order number, package type, stock item name, tax rate, and order description. Additionally, it tracks the contact person and sales person associated with each order, as well as the individuals responsible for...

AI generated content may be incorrect

Discard Apply

Generated description

User can accept this by clicking "Apply"

The user can also edit the description before accepting it



Source: Atlan

Copyright © Intelligent Business Strategies 1992-2024

32



Generative AI And Data Catalogs - Automated Metadata Enrichment Using Atlan AI To Auto-Generate Column Descriptions And Business Data Names

The screenshot shows the Atlan AI interface for a table named 'FCT_ORDERS'. A red box highlights the 'ASK ATLAN AI' dropdown menu, which contains two options: 'Description' (Autofill all columns description) and 'Business names' (Autofill all columns business names). The table's metadata, including columns and data types, is visible in the background.



Data catalog

Source: Atlan

Copyright © Intelligent Business Strategies 1992-2024

33



Atlan AI Activity Shows A Log Of The Column Description Changes Made By A User Using Atlan AI

The screenshot shows the 'Activity' log in the Atlan AI interface. It lists several updates to column descriptions:

- 'Description was updated' for the 'STOCK_ITEM_NAME' column in the 'FCT_ORDERS' table.
- 'Description was updated' for the 'STOCK_ITEM_ID' column in the 'FCT_ORDERS' table.
- 'Description was updated' for the 'FCT_ORDERS' table in the 'ANALYTICS' database.

 Red arrows point from the text annotations on the right to the corresponding activity log entries.



Data catalog

Accepted AI-generated column descriptions are immediately visible

The activity log shows they were updated by a user using Atlan AI

Source: Atlan

Copyright © Intelligent Business Strategies 1992-2024

34

Generative AI In Data Catalogs - Automated Metadata Enrichment Using AI-Generated Column Descriptions In Collibra

The screenshot displays the Collibra interface for a table asset. A modal window titled "Collibra AI Suggestions for Column Descriptions" is open, showing a list of columns with their names and AI-generated descriptions. The descriptions are placeholder text like "Lorem Ipsum is simply dummy text of the printing and typesetting industry." The modal also includes a "Change Prompt For All" dropdown, "Show Prompts" buttons for each row, and "Cancel" and "Apply" buttons at the bottom.

Source: Collibra

Copyright © Intelligent Business Strategies 1992-2024

35

Generative AI In Data Catalogs – Automatic AI-Generated Metadata Enrichment At Scale During Source Data Discovery In Atlan

The screenshot shows the Atlan "Workflow center" for "Snowflake Assets". Under the "Auto enrich using AI" section, there are checkboxes for "Description", "Readme", "Business names", and "Synthetic data". The "Description" checkbox is checked and highlighted with a red arrow and text: "Select what metadata you want automatically generated". Below this, there is a section for "Atlan AI Zero touch AI enrichment" which is also highlighted with a red box. A text box at the bottom of the screenshot states: "As Snowflake data assets are automatically discovered in Atlan, generative AI can be invoked to auto-enrich the metadata".

Source: Atlan

Copyright © Intelligent Business Strategies 1992-2024

36

Several Vendors Support Predefined Classifiers To Classify Sensitive Data Types – E.g. 200+ Pre-Defined ML Model Classifiers Used In Microsoft Purview

Examples

- People's names
- Phone numbers
- Email addresses
- Postal codes
- Passport numbers
- Driving license numbers
- Social security numbers
- Credit card numbers
- Bank account numbers
-

These are examples of very specific sensitive data classifications

You can also add your own sensitive data types

Data

Scan & classify

You need to identify where this sensitive data is in all data stores and content

Content

Scan & classify

Copyright © Intelligent Business Strategies 1992-2024

Data Profiling On-Premise And Cloud Master Data Sources

SBV = Shared Business Vocabulary

Copyright © Intelligent Business Strategies 1992-2024

Automated Data Quality Profiling – IBM Cloud Pak For Data & IBM Knowledge Catalog

The screenshot displays the IBM Cloud Pak For Data interface for a dataset named 'BANK_CUSTOMERS'. On the left, a sidebar shows the 'Data quality' section with a 'Data quality score' of 95% and a '-2%' change. Below this, a table lists columns and their quality scores: CUSTO... (100%), NAME (100%), ADDRE... (100%), ZIP (78%), CREDIT... (100%), AGE (99%), and GENDER (100%). The main area shows a table of 16 columns with their analysis status (all 'Completed'), last analyzed date ('last month'), data class, term, format, nullability, uniqueness, minimum, and maximum values. A red box highlights the 'Publish' button in the top right, with a text overlay: 'Publish puts the DQ metadata into Watson Knowledge catalog for all to see'. The source is cited as IBM.

Name	Score	Delta
CUSTO...	100%	0%
NAME	100%	0%
ADDRE...	100%	0%
ZIP	78%	0%
CREDIT...	100%	0%
AGE	99%	0%
GENDER	100%	0%

Source: IBM

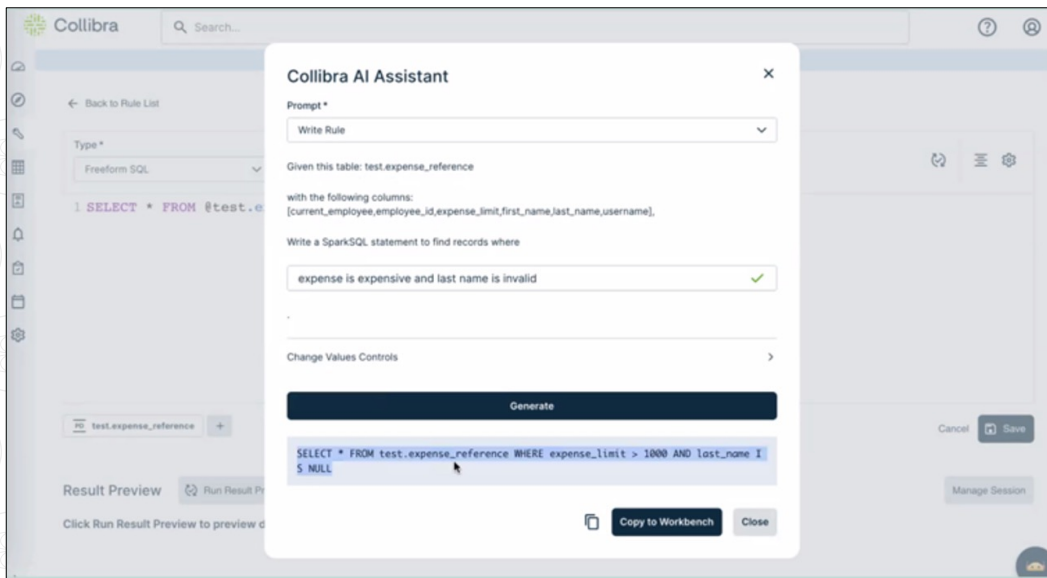
Informatica Allows Data Quality Rules To Be Defined Using Natural Language And Its CLAIRE AI Engine Will Recommend A Data Quality Rule

The screenshot shows the Informatica Data Governance and Catalog interface. A dialog box titled 'Select Technical Rule Reference' is open, allowing users to create a new rule or pick an existing one. The 'Describe the rule' field contains the text 'ID cannot be null'. The 'CLAIRE Recommendations' section shows a recommendation: 'IF C_ID != NULL THEN TRUE'. The dialog also displays the rule name 'BT_C_ID_C... Feb 04 2022 173856' and the description: 'This is a CLAIRE generated rule based on the description "ID cannot be null"'. The input field is set to 'C_ID'. The source is cited as Informatica.

Source: Informatica



Generative AI in Data Governance – Defining Data Quality Data Validation Rules Using Collibra AI Assistant



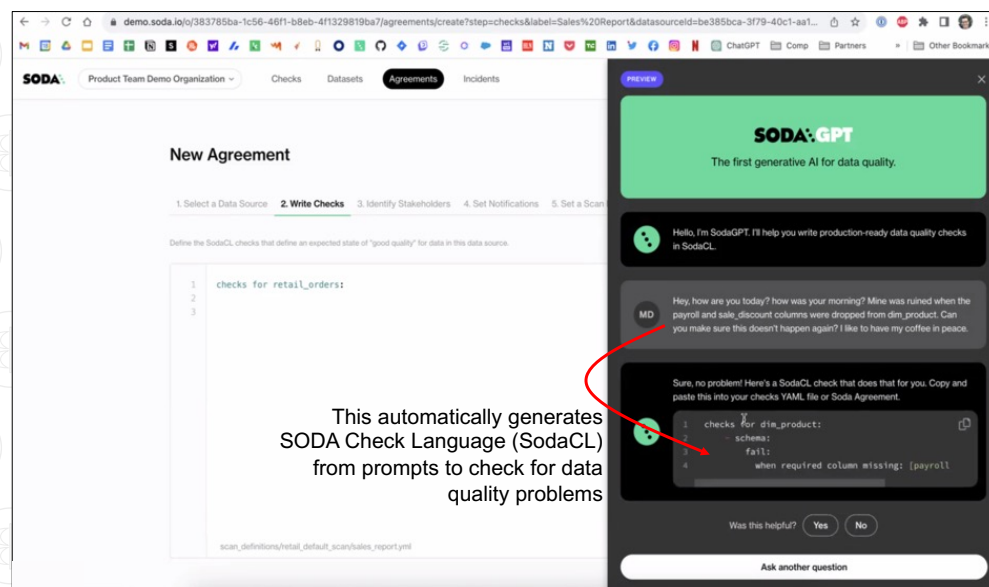
Source: Collibra

Copyright © Intelligent Business Strategies 1992-2024

41



Generative AI In Data Governance – SodaGPT Lets Business Analysts Create Data Quality Checks In Natural Language That Can Be Shared With Data Engineers To Fix Data



This automatically generates Soda Check Language (SodaCL) from prompts to check for data quality problems

SodaCL is a low code abstract language built on SQL and Spark



Copyright © Intelligent Business Strategies 1992-2024

42

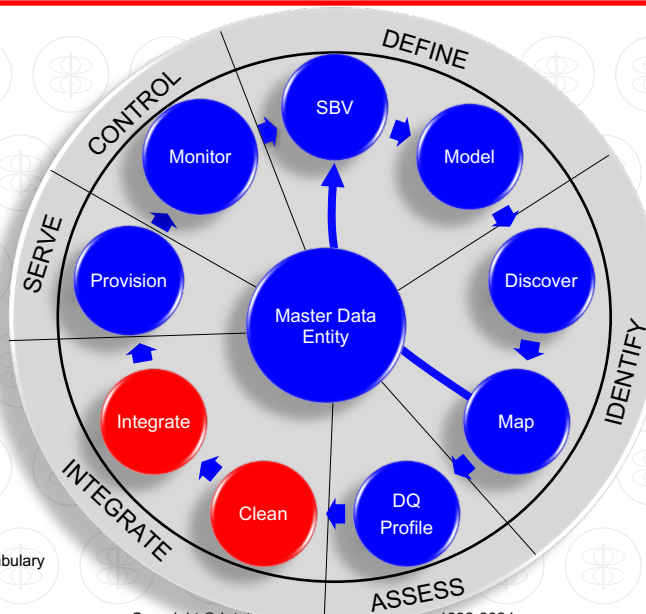


Generative AI In Data Governance – It Is Also Possible To Automatically Extract And Generate Data Governance Rules From Documents, E.g., IBM Knowledge Catalog

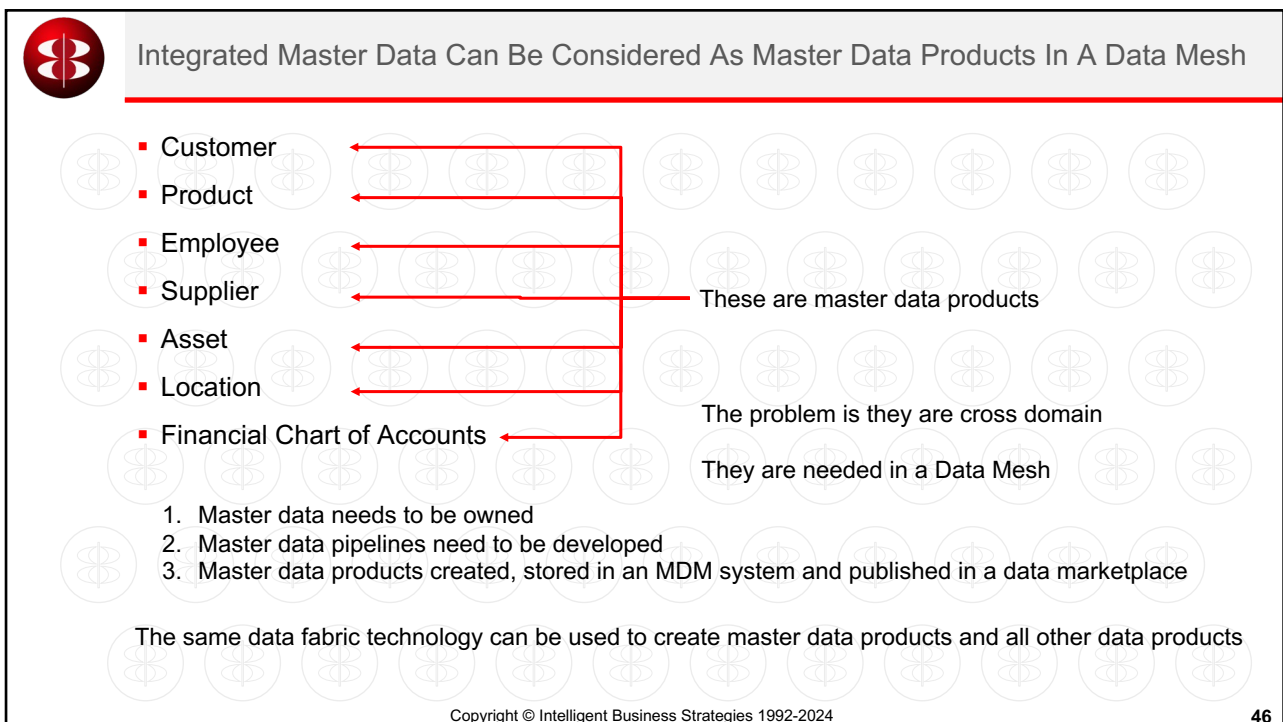
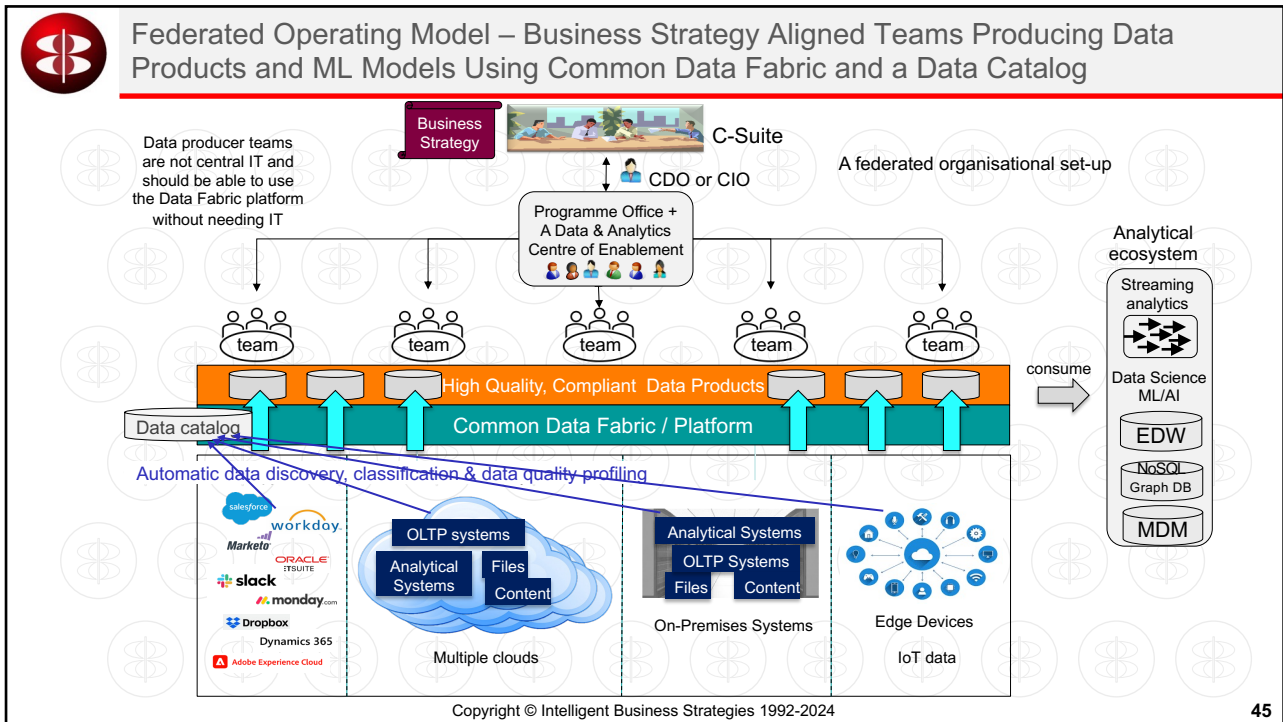
The screenshot shows a web browser displaying the IBM Watson Knowledge Catalog interface. On the left, a document titled "MONEY LAUNDERING RED FLAGS WIRE TRANSFERS" is open, listing several red flags such as "Wire transfer to bank secrecy haven countries" and "Incoming/Outgoing wire transfers with instructions to pay upon proper identification". On the right, a "Done" panel shows "12 RULES EXTRACTED" and "2 TAGS CREATED". Below this, a list of extracted rules is visible, including "Select all", "cash deposit may", "could involve wire transfers", and "customer may receive many small incoming wires".



Where Are We? – Data Cleansing And Integration



SBV = Shared Business Vocabulary



Prompt-Based Search and AI-Assisted Question Generation in Data Catalogs - Product Example: Data.world

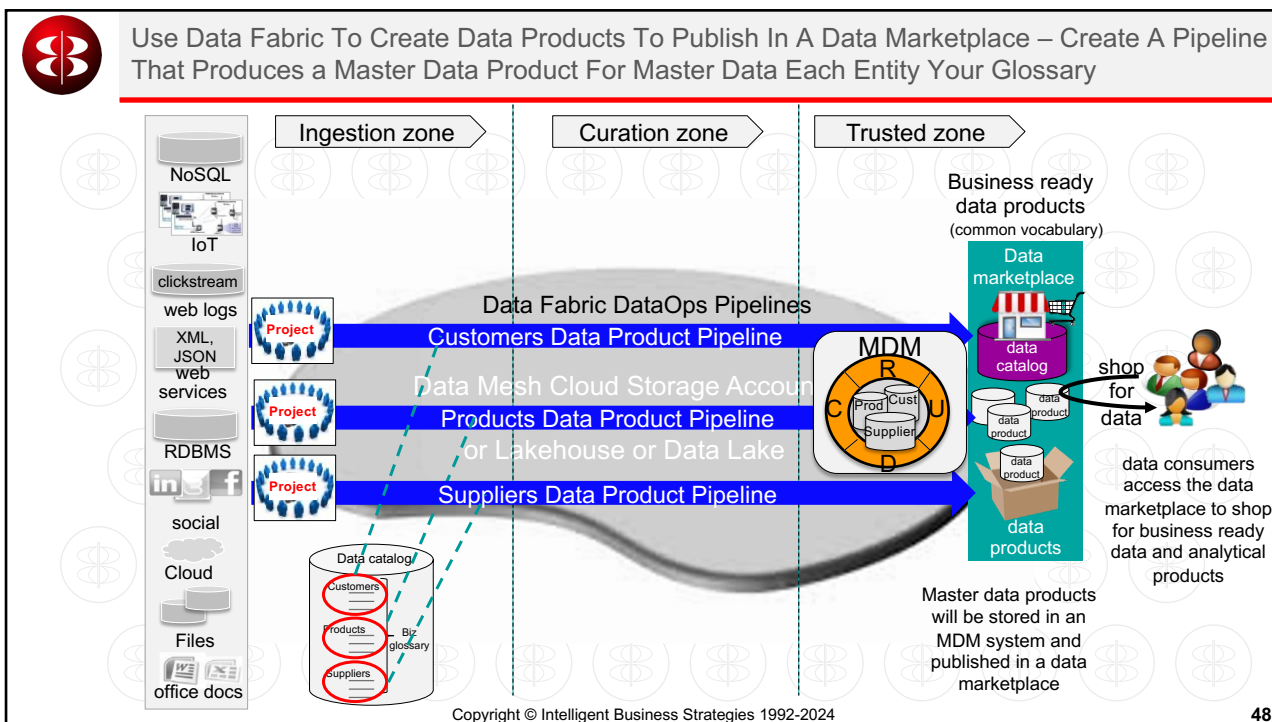
AI-Assisted Search

Source: data.world

Automated AI-Assisted Question Generation

data.world now has generative AI bots called Eureka bots

Copyright © Intelligent Business Strategies 1992-2024 47





Generative AI In Data Engineering – SnapLogic SnapGPT Prompt Based Data Engineering

Data engineering



The screenshot shows the SnapLogic Studio interface. A 'SnapGPT Preview' dialog is open, displaying a pipeline with steps: File Reader, CSV Parser, Mapper, Filter, CSV Formatter, and File Writer. A red box highlights the 'Import on new tab' button. A red arrow points from the 'Import on new tab' button to the 'Generated pipeline preview' text. To the right, a SnapGPT chat window shows a prompt: 'Read a csv file, then map Registered, Last Name, Email and % complete, then filter records where Registered is equal to or greater than 2023-05-15, and % complete is equal to or less than 40, then write records to csv file'. Below the chat window are buttons for 'Generate a pipeline' and 'Describe a pipeline'.

Copyright © Intelligent Business Strategies 1992-2024



Generative AI Prompt Based Data Engineering – SnapLogic SnapGPT Pipeline Configuration Wizard

Data engineering



The screenshot shows the SnapLogic Studio interface with the 'Pipeline Configuration Wizard' open. The wizard is in 'Step 1 of 3' and is configuring settings for a 'File Reader' snap. The configuration options include: File* (dropdown), Prevent URL encoding (checkbox), Enable staging (checkbox), Number of retries (input field), and Retry interval (seconds) (input field). There is also an 'Advanced properties' section with a table for Properties and Values, and a 'Snap Execution' dropdown. A 'Next' button is visible. To the right, a SnapGPT chat window shows a prompt: 'Read a csv file, then map Registered, Last Name, Email and % complete, then filter records where Registered is equal to or greater than 2023-05-15, and % complete is equal to or less than 40, then write records to csv file'. Below the chat window are buttons for 'Generate a pipeline' and 'Describe a pipeline'.

Source: SnapLogic

Copyright © Intelligent Business Strategies 1992-2024



Data Cleansing Is Important Before Attempting A Match – Product Example: CluedIn Cluster And Edit Uses ML To Find Similar Data To Standardise

The screenshot shows the CluedIn interface for a data cleansing project. It displays a table with 98 rows. A cluster analysis for the 'employee.country' column is shown, identifying two groups of similar values: 'United States' (27 rows) and 'United States of America' (4 rows), and 'Brazil' (46 rows) and 'Brasil' (5 rows). The interface offers options to merge these values and provides a 'Cluster & Edit' tool to standardize the data.

Source: CluedIn

Copyright © Intelligent Business Strategies 1992-2024

51



Automatic Inferred Lineage Discovery Using Graph Analytics – Hitachi Vantara Lumada Data Catalog Suggested (Recommended) Joins

The screenshot shows the Lumada Data Catalog interface. It displays a lineage graph for a resource named 'Campaign_A4.csv'. The graph shows the flow of data from upstream sources like 'Campaign_A2003_rh' and 'Campaign_A2003.csv' through 'Campaign_DW' to the target resource 'Campaign_A4.csv'. A sidebar on the right shows 'Suggested Operation (SUGGESTED)' with a 'JOIN' action and buttons for '+ ADD SOURCE', 'ACCEPT LINEAGE', 'REJECT LINEAGE', and 'VIEW MAPPING'.

52

Guided Data Cleansing and Transformation - SAS Data Studio Offers Suggestions (Recommendations)

The screenshot shows the SAS Data Studio interface. On the left, a 'Suggestions' sidebar lists various data cleansing and transformation tasks such as 'Standardize', 'Trim whitespace', 'Center and scale', and 'Convert column'. The main workspace displays a data table for 'MYCREDITSMS' with columns: Name, Credit, Expect, SMS, State, Age, and Income. Below the table, a '1. Join' step is visible in the plan view. The interface includes a top navigation bar and a right-hand panel for plan management.

Reltio ML-Powered MatchIQ - Automate ML Driven Matching and Merge Based on Scoring

Prepare data, train and review an ML model, review results to determine accuracy, approve models

The screenshot shows the Reltio MatchIQ web interface. The interface is divided into several sections: 'External match', 'Prepare' (Models in progress of collecting data), 'Train' (Models in training phase), 'Review' (Models pending a review), and 'Approved models' (Models done). The 'Approved models' section shows two models: 'Test_4' and 'Test_4 - Imported', both with 'MANAGE PUBLISH' buttons. The interface includes a top navigation bar and a left-hand sidebar with entity types like HCP, Consumer, and Employee.

Source: Reltio

Copyright © Intelligent Business Strategies 1992-2024



Reltio ML-Powered MatchIQ - Specify ML-score Thresholds to Trigger Automated Merge Or Other Matching Actions On MatchIQ Pairs

Publish settings: Individual Match IQ 003

Label: Match IQ recommended
Name displayed in Hub (MDM UI):

Scope: Internal External Match by Operational Value (OV) only
Select to publish this model for Internal and/or External Scope. By selecting this option, this match model will only be applied to OV values.

Match action	Action label	Relevance score range
Automatically merge		0.83 ————— 1
Create potential matches	High Confidence	0.56 ————— 0.83
Create potential matches	Medium Confidence	0.4 ————— 0.56

+ ADD MORE ACTION

CANCEL SAVE

Source: Reltio

Review recommended matches in the potential match screen

Copyright © Intelligent Business Strategies 1992-2024

55



Informatica MDM – Train ML Models for AI-Driven Match and Merge

Informatica Cloud | Business 360 Console | Organization: X

https://qa-pod1-mdm.mvel.infoga.com/ui-config-app/main/be/9f9XG08m1D25fceEFZlPw4/match/ea9b371b-e57e-4574-bb16-d3134c101c16/1

Informatica Business 360 Console

Organization > Model_6ziKU

Model Configuration Declarative Rules Machine Learning Model

Training the Machine Learning Model

Training is complete and the match model is published.

30	30	0%	0%	100%
Record Pairs Reviewed	Record Pairs with Conclusive Labels	Precision	Recall	Accuracy

	Predicted as Match	Predicted as Not a Match	Total
Labeled as Match	0% (0)	0% (0)	0% (0)
Labeled as Not a Match	0% (0)	100% (9)	100% (9)
Total	0% (0)	100% (9)	

Match Fields (14)

Field Name	Field Path	Used in
Phone.Type	Phone.Communication	Declarative Rules, Training

Source: Informatica

Copyright © Intelligent Business Strategies 1992-2024

56

Batch Oriented Master Data Matching And Consolidation

Matching

- Matching is the process of comparing two records for points of similarity
- If sufficient points of similarity are found to indicate that the two records are probably duplicates of each other, then those records are flagged for consolidation
- The columns used for comparison purposes are called *match columns*
- Each match column is based on one or more columns from the base object
- Match columns are combined into *match rules* to determine the conditions under which two records are considered to be similar enough to consolidate

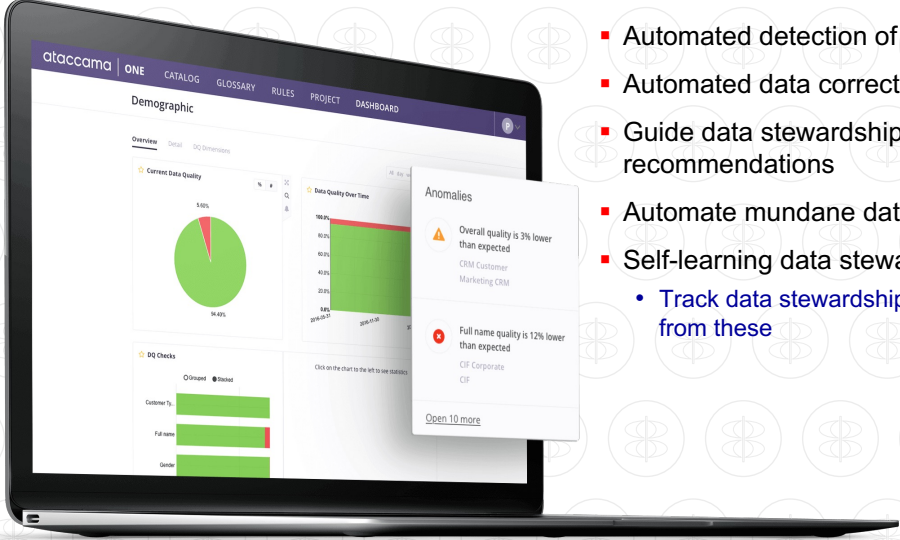
Consolidation

- After duplicate records have been identified in the match process, the consolidate process merges duplicate records into a single record

MDM System	Master ID	First Name	MN	Last Name	Address	City	State	Zip
	M-0001	Abel	Noel	Willan	161 Washington Ave.	Buffalo	NY	14263
Sales	SFA_ID	First Name	MN	Last Name	Address	City	State	Zip
	12345	Abel		Willan	161 Washington Ave.	Buffalo	NY	14263
Accounts	Cust_ID	First Name	MN	Last Name	Address	City	State	Zip
	502068	Abel	Noel	Willan	161 Washington Ave.	Buffalo	NY	14263
Marketing	Target_ID	First Name	MN	Last Name	Address	City	State	Zip
	willan05	Abel	N	Willan	Elm & Carlston Streets	Buffalo	NY	14263

Source: Informatica Copyright © Intelligent Business Strategies 1992-2024 57

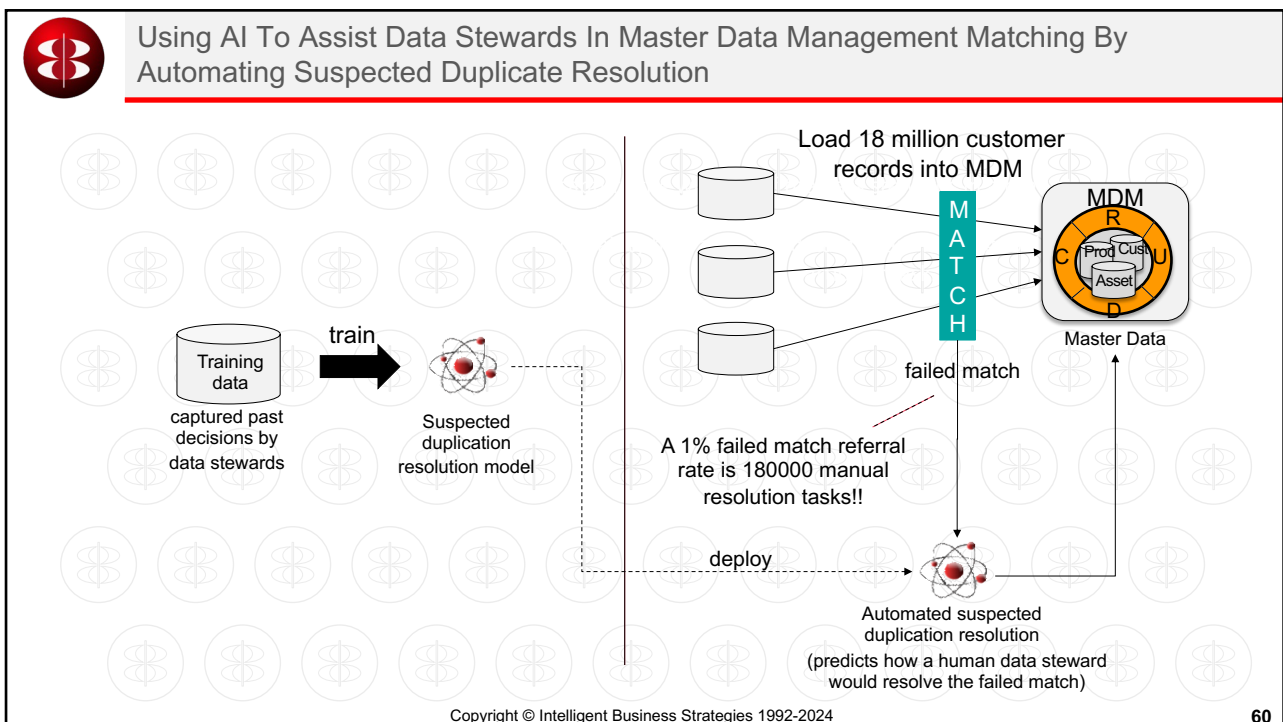
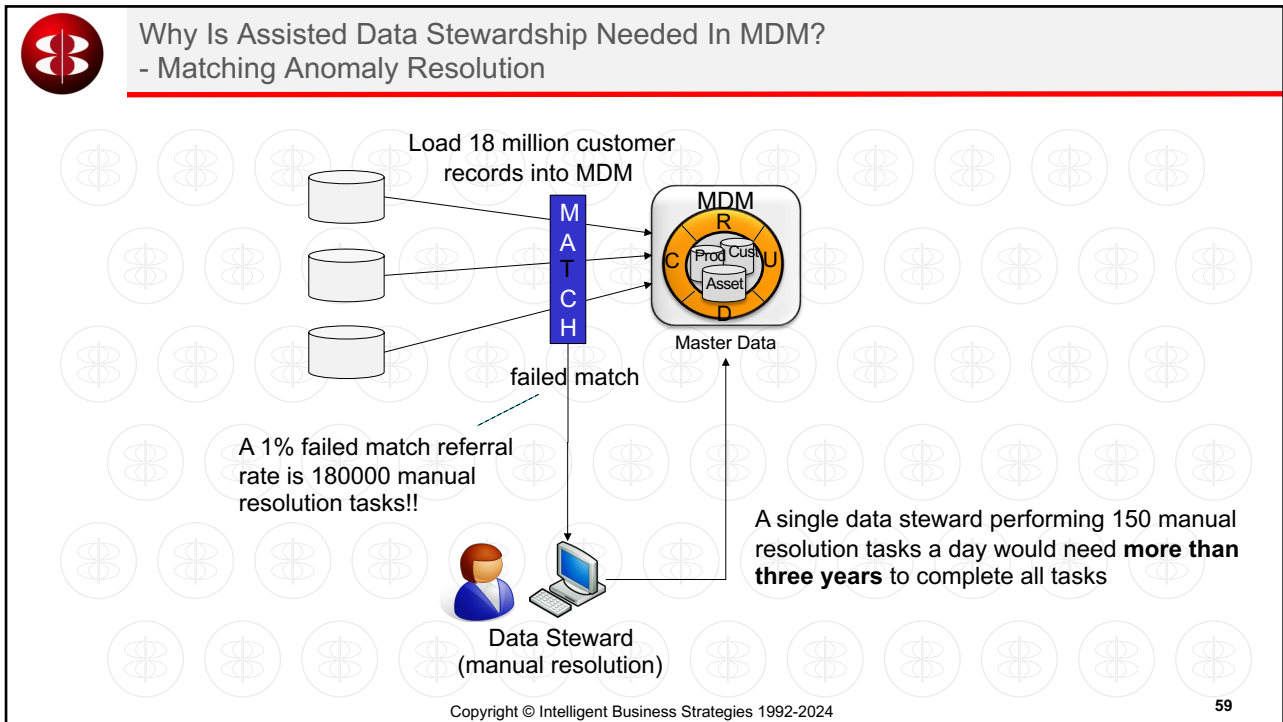
AI In Data Governance - AI Assisted Data Stewardship



The screenshot shows a dashboard with a navigation bar (ONE, CATALOG, GLOSSARY, RULES, PROJECT, DASHBOARD) and a 'Demographic' section. It features a 'Current Data Quality' pie chart, a 'Data Quality Over Time' line chart, and a 'DQ Checks' table. An 'Anomalies' pop-up window highlights two issues: 'Overall quality is 3% lower than expected' and 'Full name quality is 12% lower than expected'.

- Automated detection of poor quality data
- Automated data correction
- Guide data stewardship decisions using recommendations
- Automate mundane data stewardship tasks
- Self-learning data steward
 - Track data stewardship decisions and learn from these

Copyright © Intelligent Business Strategies 1992-2024 58



AI Assisted Data Stewardship – Phased Deployment Of AI

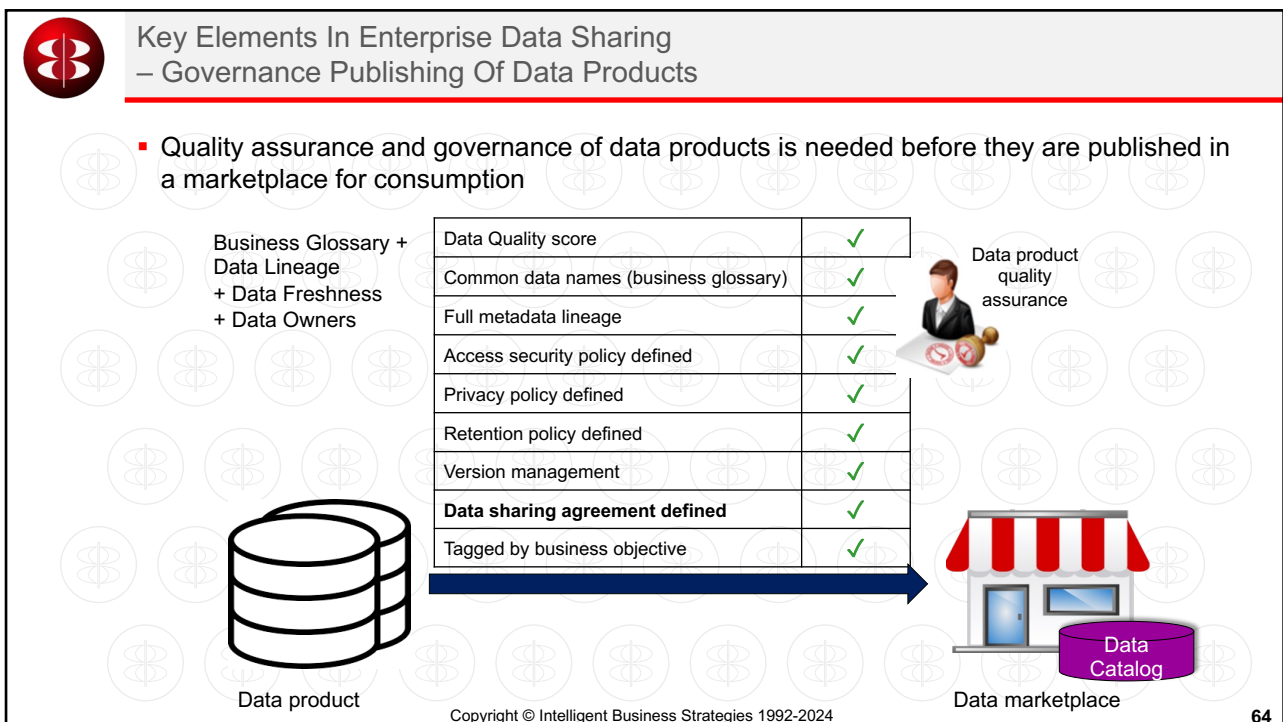
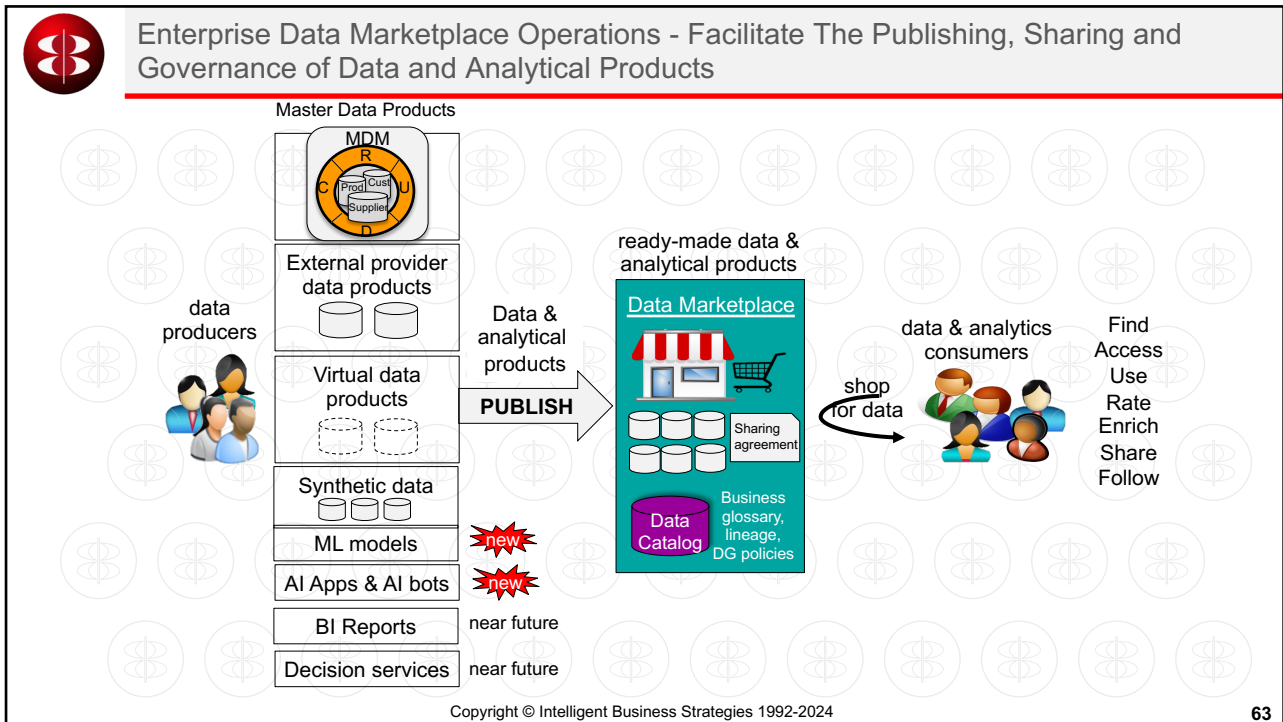
- Step 1**
 - Introduce the trained model to recommend to the data stewards how to resolve the suspected duplicate
 - Data steward simply reviews and approves
- Step 2**
 - Automate the process above a model prediction confidence level threshold (e.g. >95%) once data stewards trust it


Image source: IBM

Copyright © Intelligent Business Strategies 1992-2024 61

Where Are We? – Provisioning / Synchronising High Quality, Commonly Understood Master Data

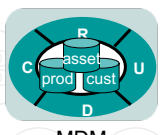
Copyright © Intelligent Business Strategies 1992-2024 62

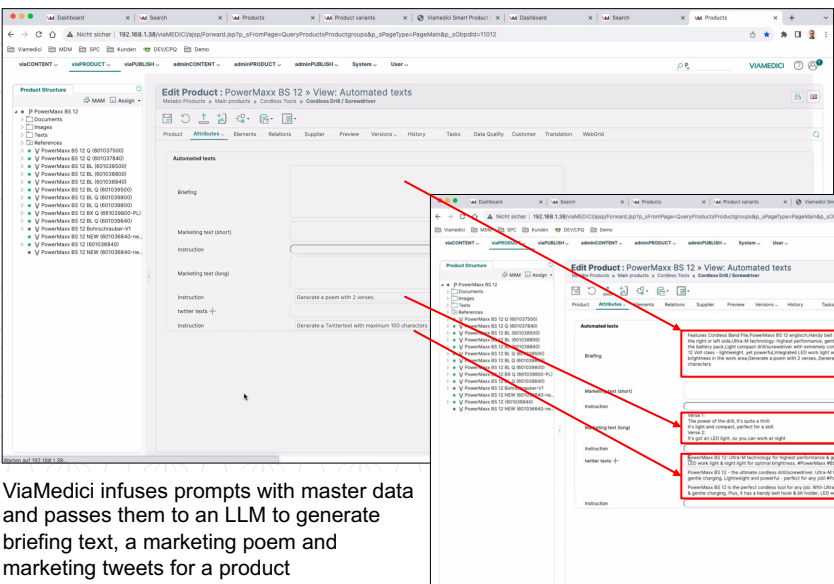




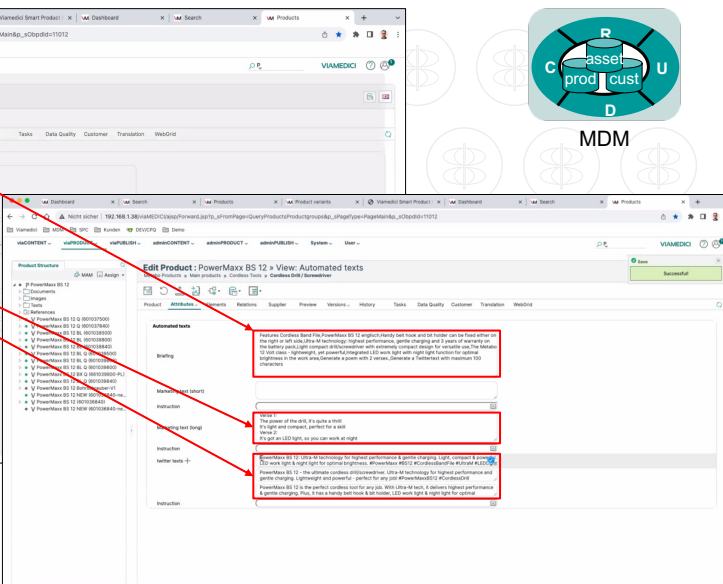
Generative AI In Master Data Management (MDM)

– Product Example: ViaMedici Product MDM & AI-Powered PIM Text Generation





ViaMedici infuses prompts with master data and passes them to an LLM to generate briefing text, a marketing poem and marketing tweets for a product



Copyright © Intelligent Business Strategies 1992-2024 65



About Mike Ferguson



 www.intelligentbusiness.biz
 mferguson@intelligentbusiness.biz
 @mikeferguson1
 (+44) 1625 520700

Mike Ferguson is Managing Director of Intelligent Business Strategies Limited. As an independent IT industry analyst and consultant, he specialises in BI / analytics and data management. With over 40 years of IT experience, Mike has consulted for dozens of companies on BI/Analytics, data strategy, technology selection, enterprise architecture, and data management. Mike is also conference chairman of Big Data LDN, the largest data and analytics conference in Europe and a member of the EDM Council CDMC Executive Advisory Board. He has spoken at events all over the world and written numerous articles. Formerly he was a principal and co-founder of Codd and Date – the inventors of the Relational Model (which caused the birth of relational databases and the SQL language), and Chief Architect at Teradata on the Teradata DBMS. He teaches popular master classes in Data Strategy, Data Warehouse Modernisation, Practical Guidelines for Implementing a Data Mesh, Big Data, How to Govern Data across a Distributed Data Landscape, Machine Learning and Advanced Analytics, and Embedded Analytics, Intelligent Apps and AI Automation.



Thank You!

Copyright © Intelligent Business Strategies 1992-2024 66



**DATA GOVERNANCE AND
MASTER DATA MANAGEMENT
CONFERENCE EUROPE**

11 - 14 March 2024 | London, UK

****Please score and comment on this session and speaker
in the event mobile app****

