Best Practices in Metadata Management

Donna Burbank
Global Data Strategy, Ltd.

October 23, 2025

DATA ARCHITECTURE STRATEGIES

Global Data Strategy
The Business of Data

Donna Burbank

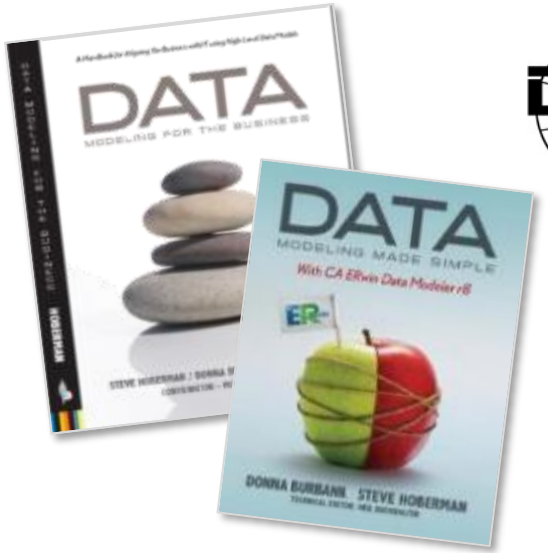
Donna is a recognized industry expert in data management with over 25 years of experience in data strategy, data governance, data modeling, metadata management, and enterprise architecture. Her background is multi-faceted across consulting, product development, product management, brand strategy, marketing, and business leadership.

She is currently the Managing Director at Global Data Strategy, Ltd., an international data management consulting company that specializes in the alignment of business drivers with data-centric technology.

In past roles, she has served in key brand strategy and product management roles for several of the leading data management products in the market.

As an active contributor to the data management community, she is a long time DAMA member, contributor to the DMBOK 2.0 and 3.0, Past President and Advisor to the DAMA Rocky Mountain chapter, and was awarded the Excellence in Data Management Award from DAMA International.

She has worked with dozens of Fortune 500 companies worldwide in the Americas, Europe, Asia, and Africa and speaks regularly at industry conferences. She has co-authored several books and is a regular contributor to industry publications. She can be reached at donna.burbank@globaldatastrategy.com
Donna is based in Boulder, Colorado, US.



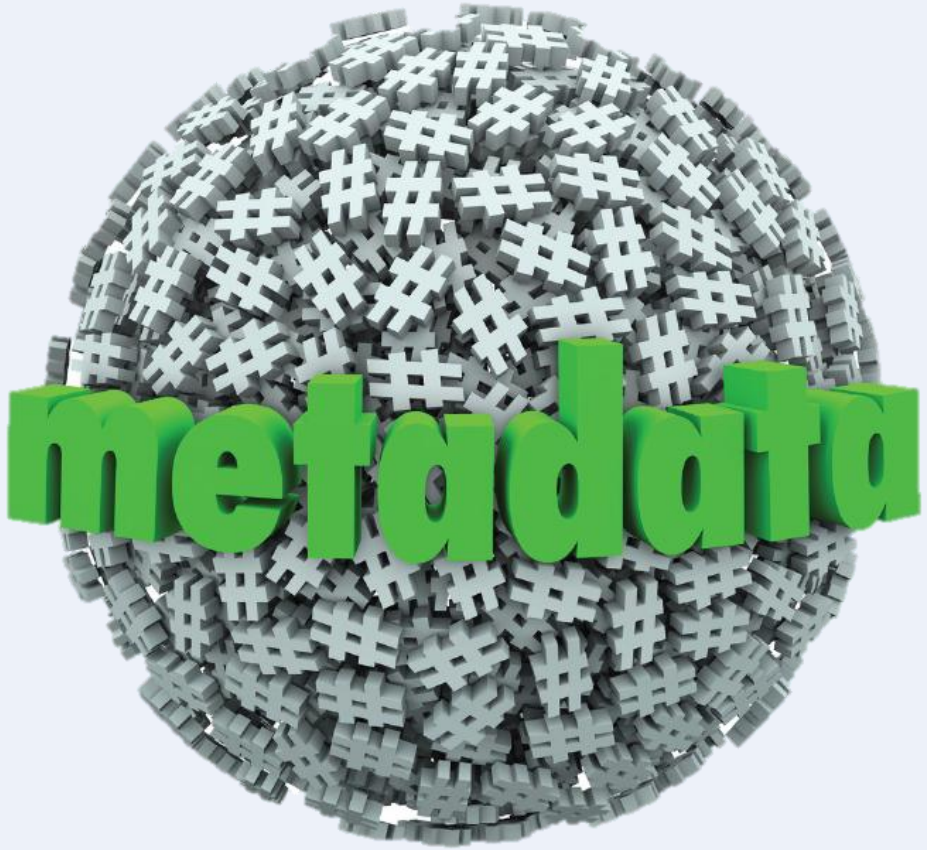
DATAVERSITY Data Architecture Strategies

This Year's Lineup

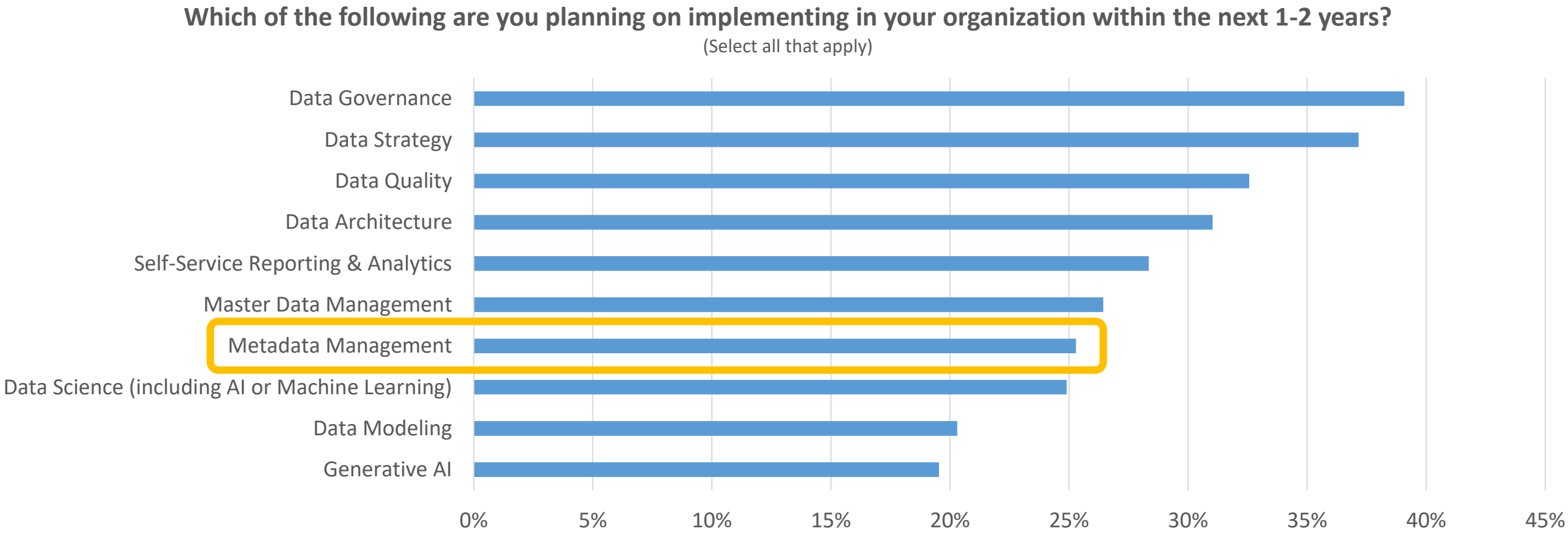
- **January** Trends in Data Architecture
- **February** Building a Data Strategy - Practical Steps for Aligning with Business Goals
- **March** Building the Right Architecture for Analytics & Reporting
- **April** Data Architect vs. Data Engineer vs. Data Scientist – Making Sense of Roles in Today's Data-Centric Organization
- **May** Master Data Management - Aligning Data, Process, and Governance
- **June** Where Data Models Fit in Today's Modern Data Architecture
- **July** Data Architecture vs. Enterprise Architecture
- **August** Data Quality Best Practices (with guest Nigel Turner)
- **September** Modern Data Architecture: Practical Options for Today's Data-Driven Organization
- **October** Best Practices in Metadata Management
- **December** The Business Value of Data Modeling

What We'll Cover Today

- **Metadata is hotter than ever**, according to several recent DATAVERSITY surveys.
- More and more organizations are realizing that **in order to drive business value from data, robust metadata is needed** to gain the necessary context and lineage around key data assets.
- While metadata has been managed for decades, **new strategies and approaches have been developed to support the ever-evolving data landscape** and provide more innovative ways to drive business value from metadata.
- **This webinar will provide an overview of metadata strategies and technologies available to today's organization** and provide insights into building successful business strategies for metadata adoption.

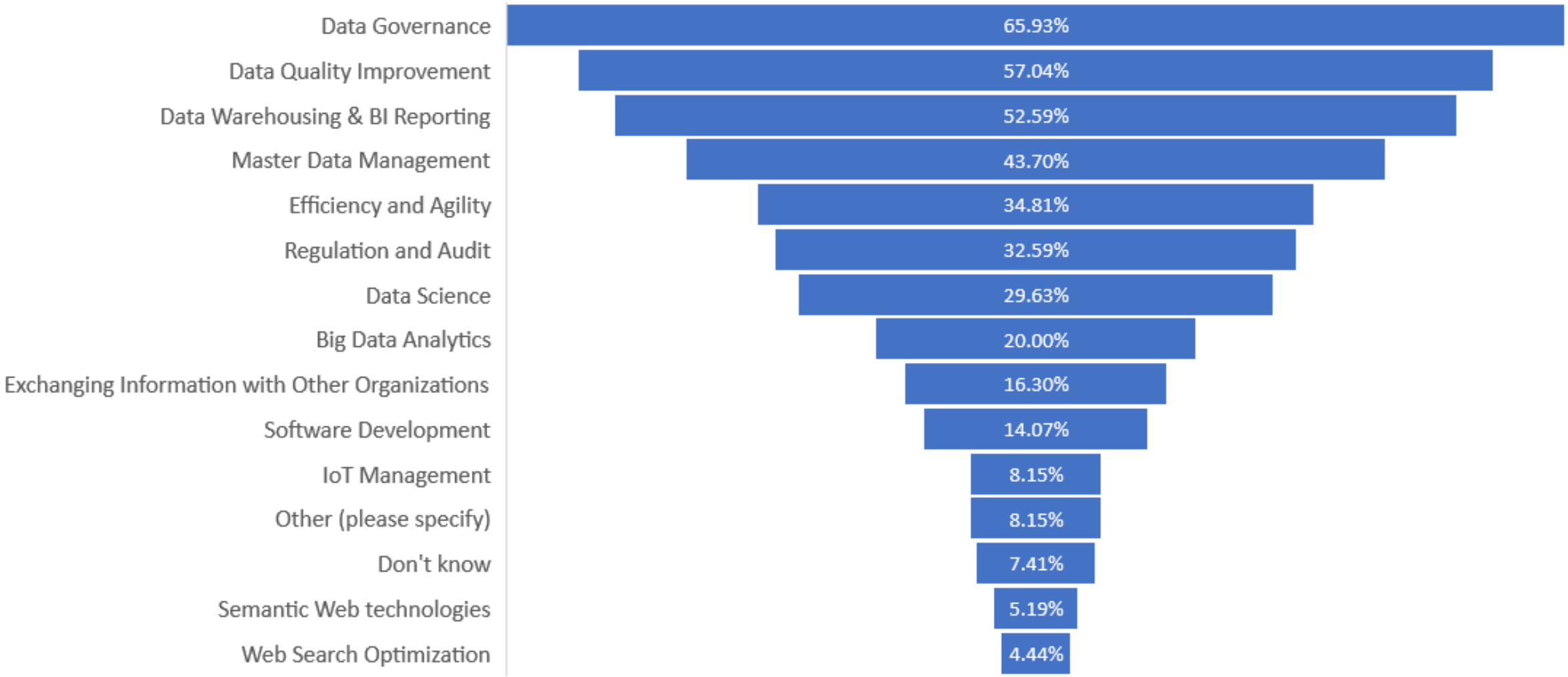


Top 10 Priorities for Coming 1-2 Years



From Trends in Data Management 2024, a DATAVERSITY Report, by Donna Burbank and Michelle Knight

Use Cases for Metadata Management



From Trends in Data Management, 2022, DATAVERSITY, by Donna Burbank and Keith Foote

Metadata Management is Part of a Wider Data Strategy

A Successful Data Strategy links Business Goals with Technology Solutions

DATA ARCHITECTURE STRATEGIES

"Top-Down" alignment with business priorities

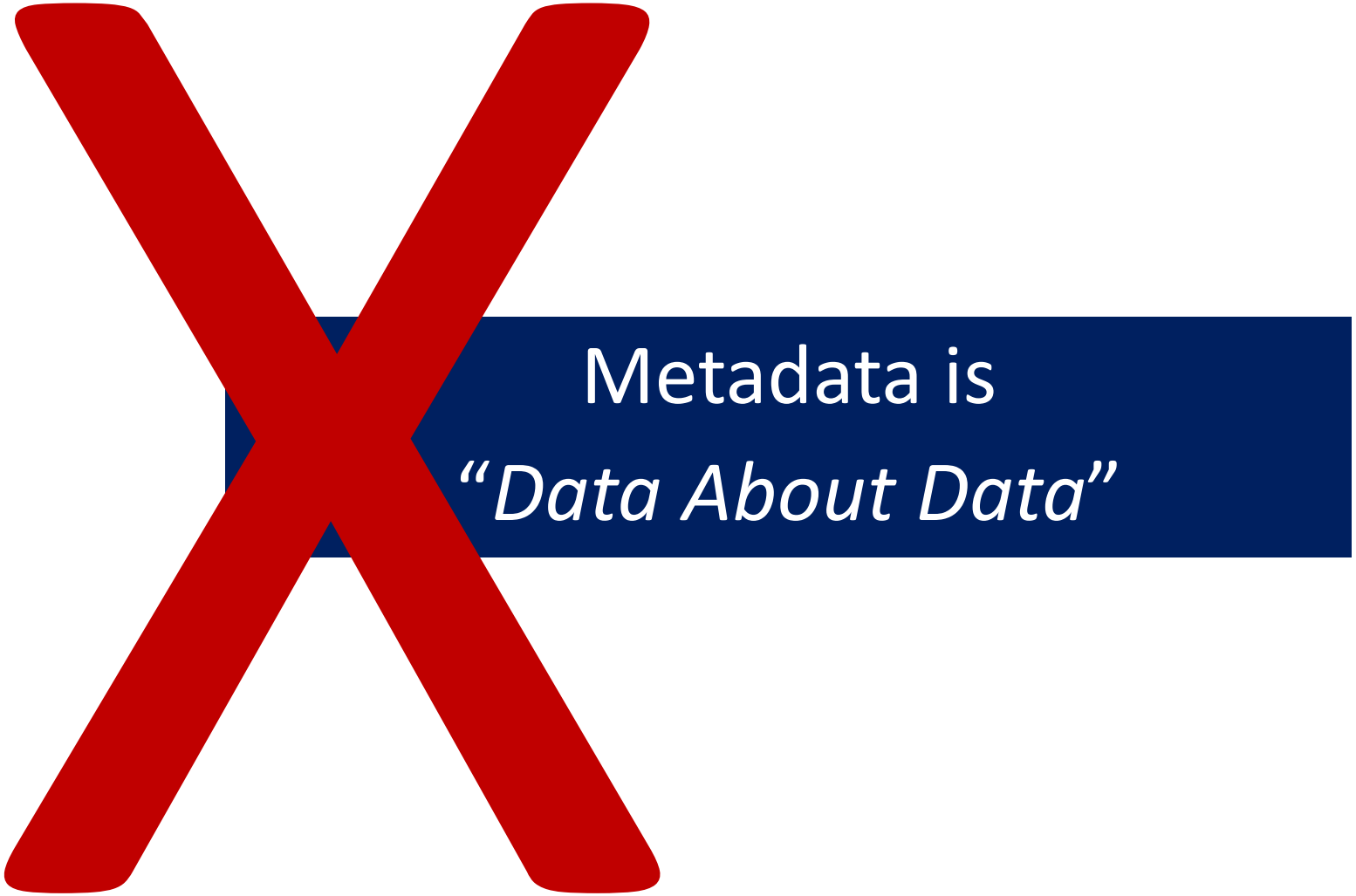
Managing the people, process, policies & culture around data

Leveraging & managing data for strategic advantage

Coordinating & integrating disparate data sources

"Bottom-Up" management & inventory of data sources

What is Metadata?



Metadata is

"Data About Data"

What is Metadata?

Metadata is Data In Context

Metadata is the "Who, What, Where, Why, When & How" of Data

DATA ARCHITECTURE STRATEGIES

Who	What	Where	Why	When	How
Who created this data?	What is the business definition of this data element?	Where is this data stored?	Why are we storing this data?	When was this data created?	How is this data formatted? (character, numeric, etc.)
Who is the Steward of this data?	What are the business rules for this data?	Where did this data come from?	What is its usage & purpose?	When was this data last updated?	How many databases or data sources store this data?
Who is using this data?	What is the security level or privacy level of this data?	Where is this data used & shared?	What are the business drivers for using this data?	How long should it be stored?	
Who "owns" this data?	What is the abbreviation or acronym for this data element?	Where is the backup for this data?		When does it need to be purged/deleted?	
Who is regulating or auditing this data?	What are the technical naming standards for database implementation?	Are there regional privacy or security policies that regulate this data?			

Data vs. Metadata

Customer

First Name	Last Name	Company	City	Year Purchased
Joe	Smith	Komputers R Us	New York	1970
Mary	Jones	The Lord's Store	London	1999
Proful	Bishwal	The Lady's Store	Mumbai	1998
Ming	Lee	My Favorite Store	Beijing	2001

Metadata

Data

Data vs. Metadata

Customer

STR01	STR02	TXT123	TXT127	DT01
Joe	Smith	Komputers R Us	New York	1970
Mary	Jones	The Lord's Store	London	1999
Proful	Bishwal	The Lady's Store	Mumbai	1998
Ming	Lee	My Favorite Store	Beijing	2001

Metadata?

Data

Metadata adds Context & Definition

Customer

First Name	Last Name	Company	City	Year Purchased
Joe	Smith	Komputers R Us	New York	1970
Mary	Jones	The Lord's Store	London	1999
Proful	Bishwal	The Lady's Store	Mumbai	1998
Ming	Lee	My Favorite Store	Beijing	2001

Is this the city where the customer lives or where the store is located?

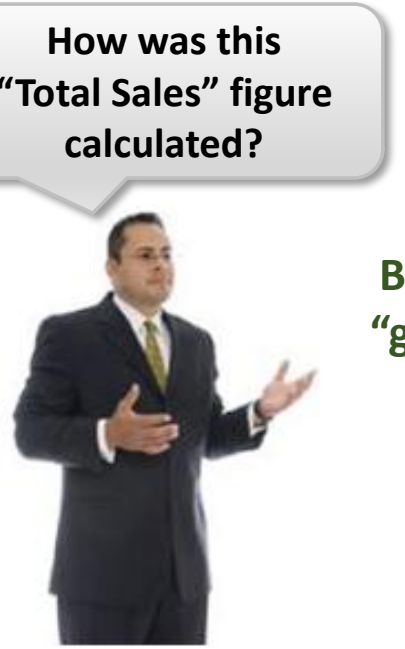
Definition	Last Name represents the surname or family name of an individual.
Business Rules	In the Chinese market, family name is listed first in salutations.
Format	VARCHAR(30)
Abbreviation	LNAME
Required	YES
Etc.	Numerous technical & business metadata including security, privacy, nullability, primary key, etc.

Metadata is Needed by Business Stakeholders

Making business decisions on accurate and well-understood data

80% of users of metadata are from the business, according to a DATAVERSITY survey[1].

"Metadata helps both IT and business users understand the data they are working with. Without Metadata, *the organization is at risk for making decision based on the wrong data*."[1]



Business users often "get" metadata more than IT does!

[1] *Emerging Trends in Metadata Management*, 2016, DATAVERSITY, by Donna Burbank and Charles Roe

Business Meaning & Context is Critical

Show me all customers by region

Businessperson

"Does this include current customers only? Or lapsed customers as well?

"How do you define region?"

"Can a customer have a billing address in more than one region?"

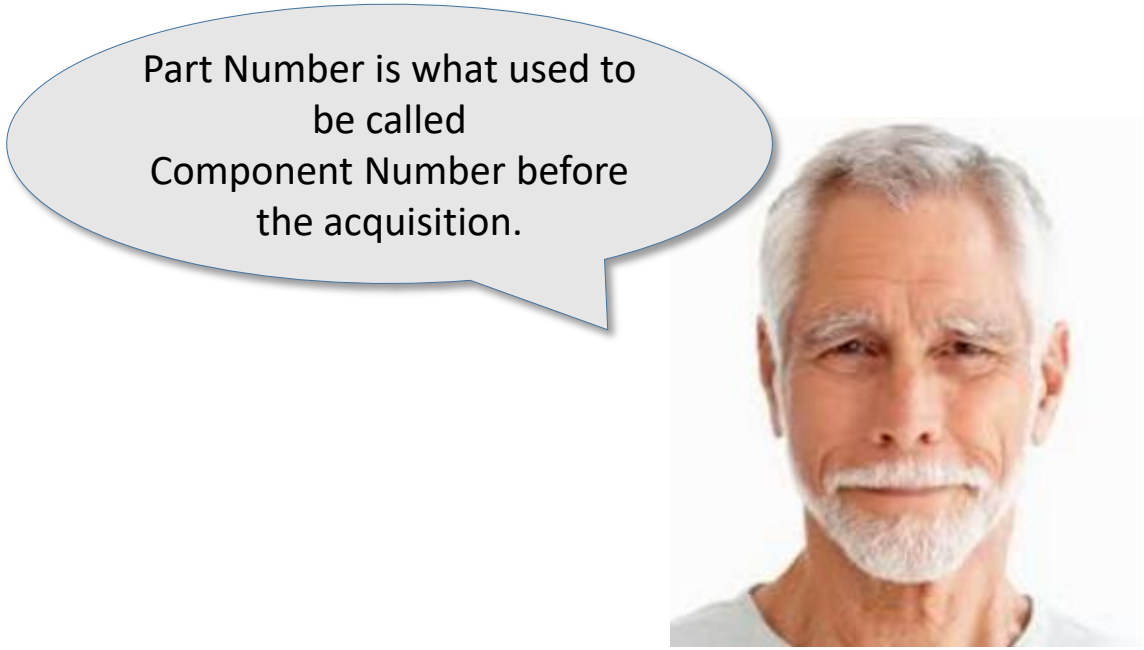
"Do we have to obfuscate PII?"

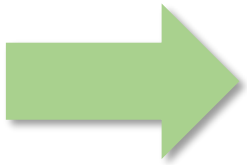
Data Architect

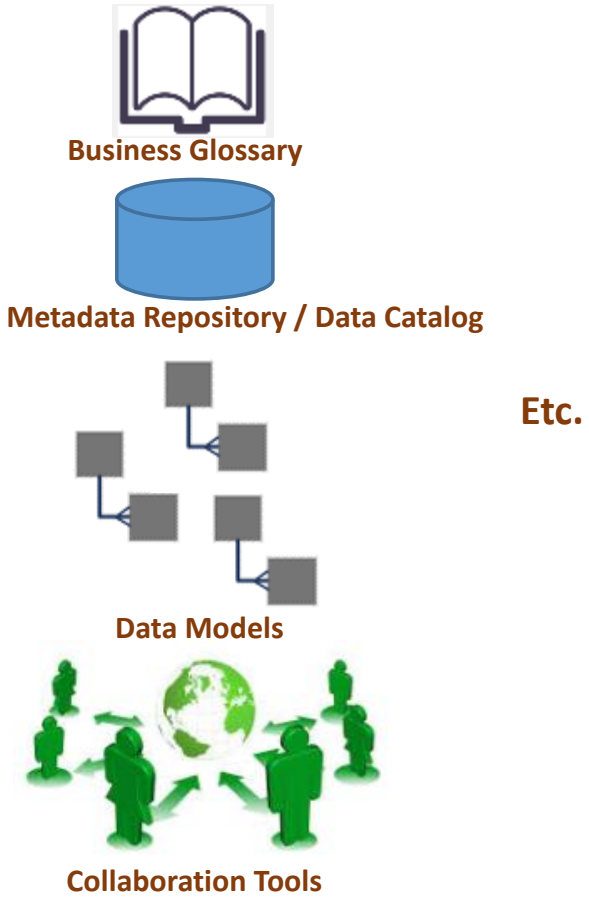
Capturing & Storing Business Metadata

Avoid the dreaded “I just know”

- Much business metadata and the history of the business exists in employee’s heads.
- It is important to capture this metadata in an electronic format for sharing with others.
- Avoid the dreaded “I just know”







Etc.

Who Uses Metadata?

Business Definitions

From Data Modeling for the Business by Hoberman, Burbank, Bradley, Technics Publications, 2009

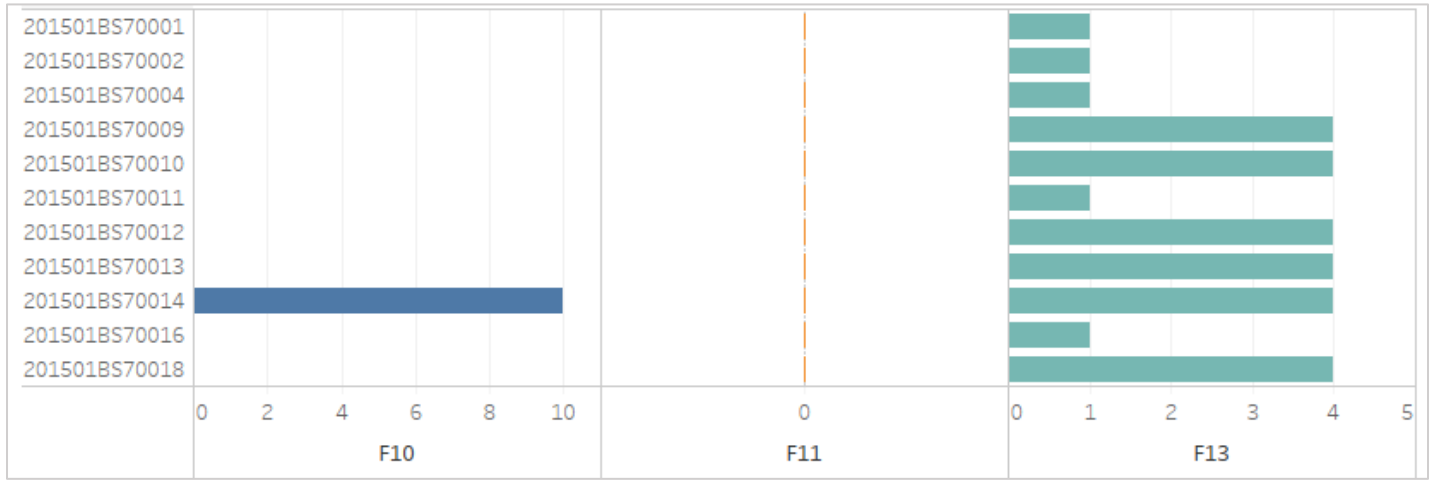
A Very Expensive Example - NASA

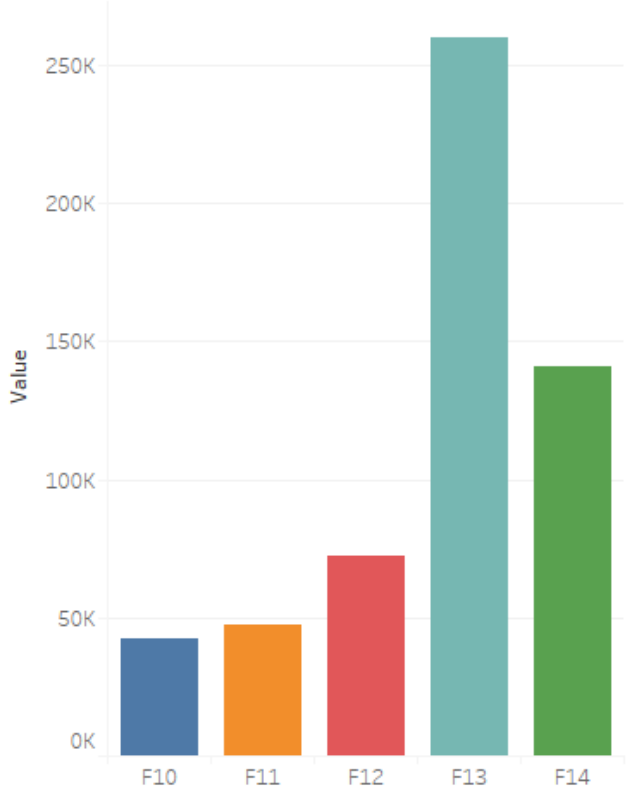
- On September 23, 1999 NASA lost the **$125 million** Mars Climate Orbiter spacecraft after a 286-day journey to Mars.
- **Missing Metadata was the culprit**
  - Thruster data was sent in English units of pound-seconds (lbf s) instead of Metric units of newton-seconds (N s)
- This metadata inconsistency caused thrusters to fire incorrectly, sending the craft off course – 60 miles in all (96.56 km).
- In addition to the financial cost of the orbiter were the additional issues of:
  - Brand and Reputational Damage
  - Lost Opportunities for research on the Martian atmosphere & climate

Data is Only as Good as the Metadata

Open Data Example: Road Safety - Vehicles by Make and Model

Abc MakeModel2015v2.csv F1	# MakeMod... F2	# MakeMod... F3	# MakeMod... F4	# MakeMod... F5	# MakeMod... F6	# MakeMod... F7	# MakeMod... F8	# MakeMod... F9	# MakeMode... F10	# MakeMode... F11	# MakeMode... F12	# MakeMode... F13	# MakeMode... F14	# MakeMode... F15
201501BS700...	2,015	1	9	0	9	0	8	0	0	0	0	1	1	6
201501BS700...	2,015	1	9	0	9	0	2	0	0	0	0	1	1	6
201501BS700...	2,015	1	3	0	18	0	8	0	0	0	0	1	1	6
201501BS700...	2,015	2	19	0	6	0	8	0	0	0	0	3	1	1
201501BS700...	2,015	1	9	0	9	0	8	0	0	0	0	4	1	6
201501BS700...	2,015	2	9	0	18	0	0	0	0	0	0	1	1	6
201501BS700...	2,015	1	9	0	13	0	8	0	0	0	0	3	1	6
201501BS700...	2,015	2	5	0	7	0	8	0	0	0	0	1	1	6

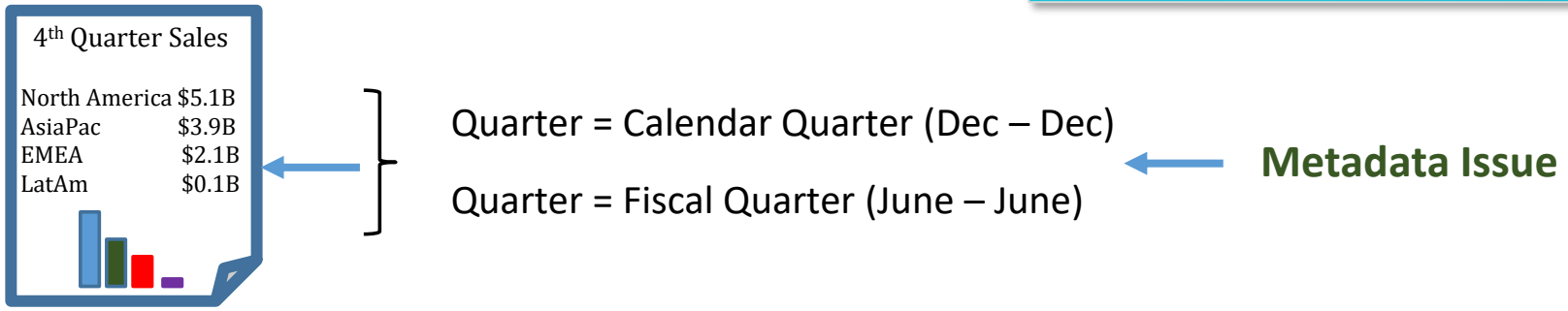




Financial Reporting – What is a Year?

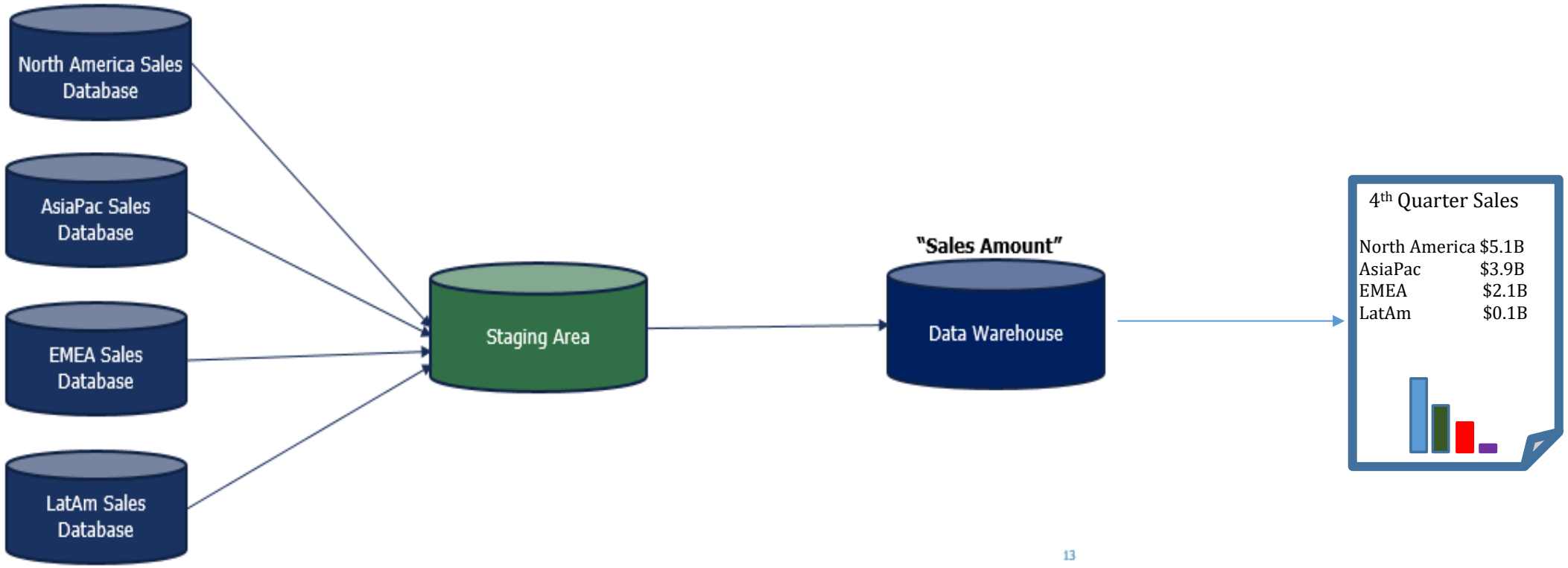
An international retail chain was comparing
4th Quarter Sales across regions.

- **Typically the 4th quarter sees a spike in revenue**, due to numerous holidays in the November & December timeframes
- But **Latin American sales from a newly-acquired subsidiary were particularly low** that quarter, prompting questions:
  - Do we need to increase marketing in that region?
  - Is this the wrong market for our products? Should we close retail stores?
- Further research determined that the **Latin American branch was using a Fiscal year of June – June,** rather than the calendar year used by the rest of the world.
- A metadata issue (mismatched definitions) caused business confusion and potentially misspent funds & effort

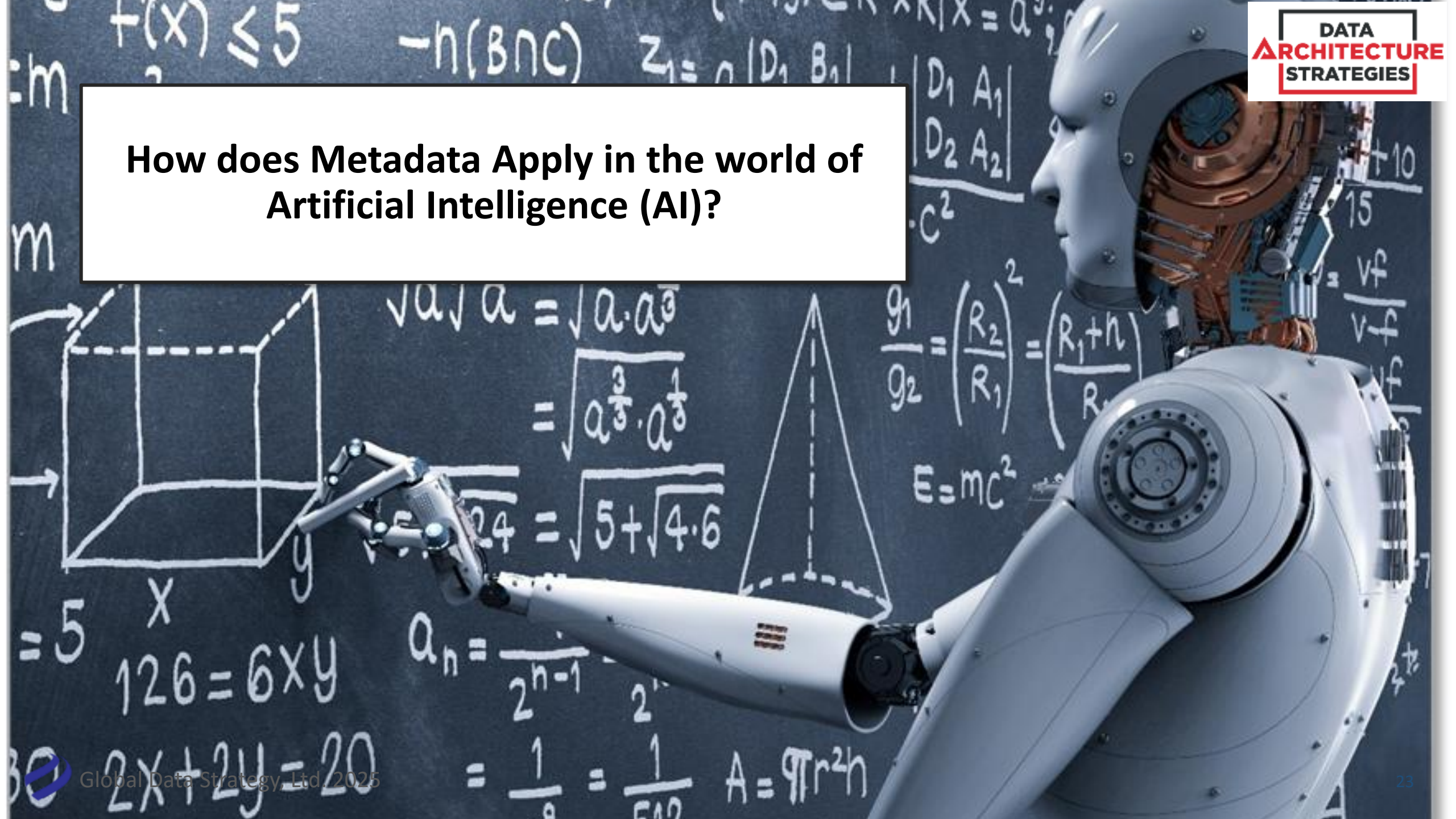


Audit & Traceability

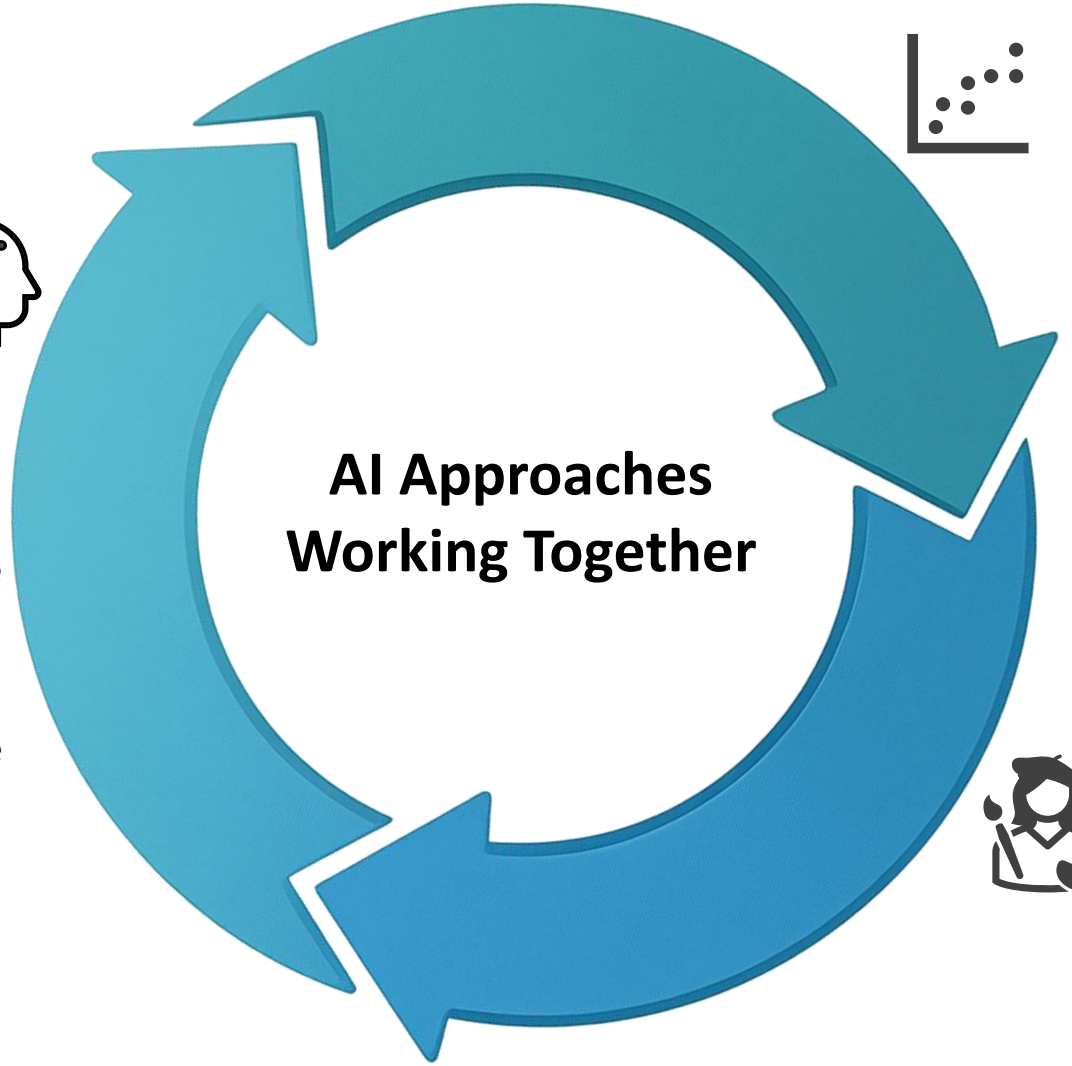
- This reporting error spurred an internal audit to evaluate **how financial figures were calculated**.
- Because this company had **good metadata tracking and lineage**, they were easily able **to show how information was sourced & manipulated to create key reports.**



How does Metadata Apply in the world of Artificial Intelligence (AI)?



Let's Take and Example: Workforce Automation with AI



Analytical AI

- Analyzes information from several sources including Inventory, Sales, & Supplier Data
- Detects Inventory shortage in a given Region

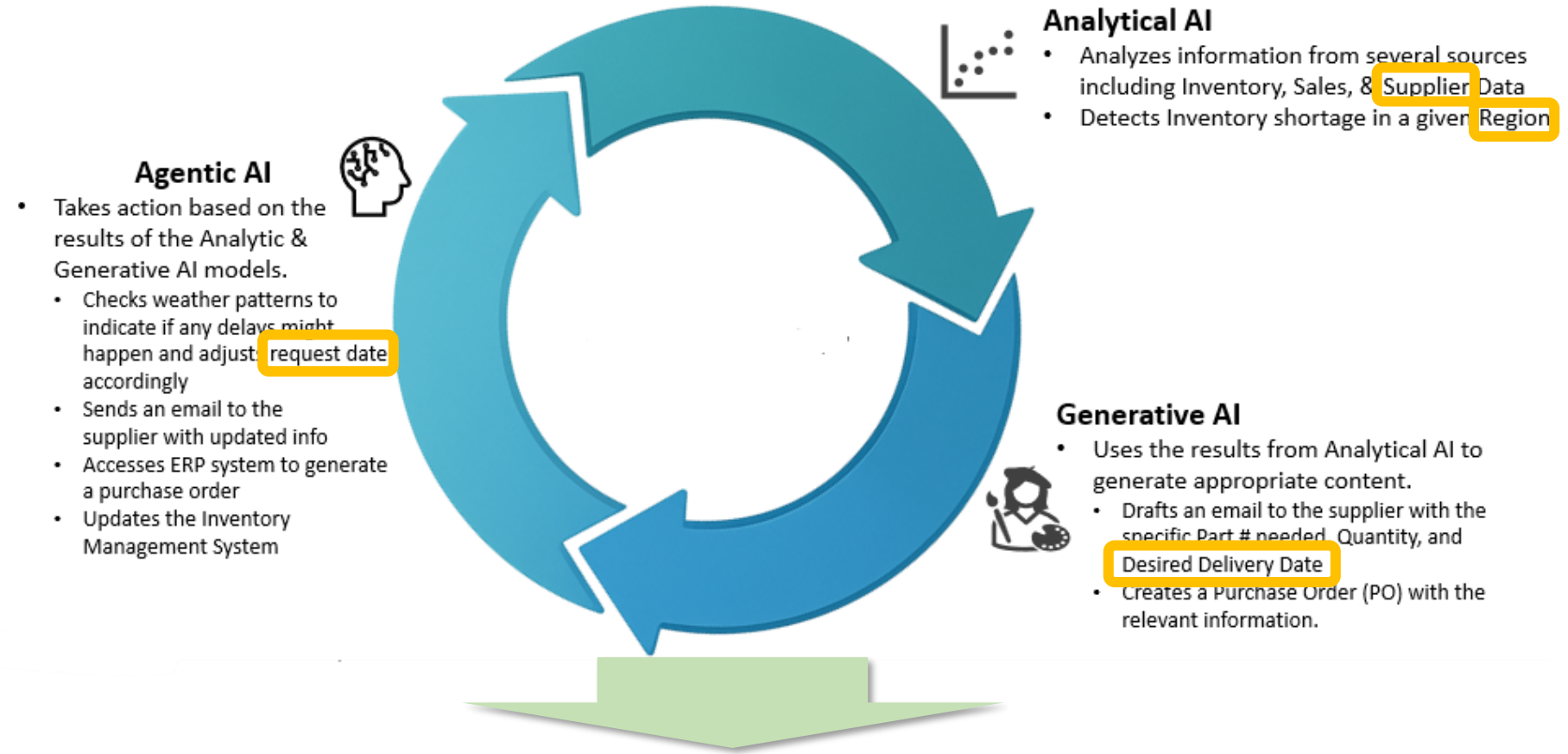
Agentic AI

- Takes action based on the results of the Analytic & Generative AI models.
  - Checks weather patterns to indicate if any delays might happen and adjusts request date accordingly
  - Sends an email to the supplier with updated info
  - Accesses ERP system to generate a purchase order
  - Updates the Inventory Management System

Generative AI

- Uses the results from Analytical AI to generate appropriate content.
  - Drafts an email to the supplier with the specific Part # needed, Quantity, and Desired Delivery Date
  - Creates a Purchase Order (PO) with the relevant information.

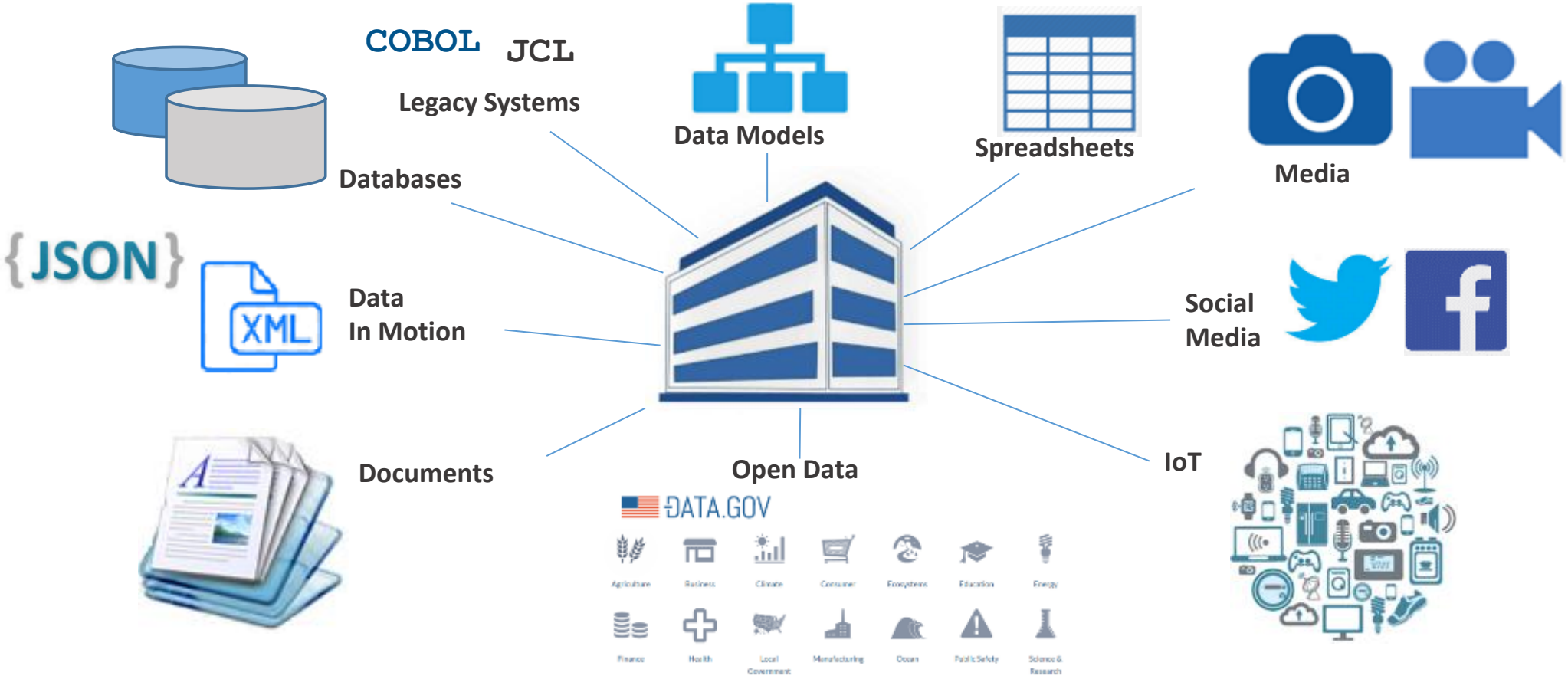
Metadata is More Important Than Ever with AI



- How do we define **Region**?
  - Sales Region?
  - Geographical Region?
- Is our **Supplier** data consistent?
  - Supplier address are stored in different fields in different systems
  - One system identifies Suppliers by DUNS number, another by Tax ID
- Are our **Dates** consistent?
  - The weather data is from Europe, the requestor is from Argentina, and the Supplier is from the US
    - Is 1/9 the beginning of January or the beginning of September?
    - Is this summer or winter?
  - Does Request Date & Delivery Date mean the same thing?

Metadata Across & Beyond the Organization

- Metadata exists in a wide variety of sources across & beyond the organization.
- In the following slides, we'll go through some examples.



COBOL Copybook Metadata

- **What is a COBOL Copybook?** – In COBOL, a copybook file is used to define data elements that can be referenced by many programs
- **What is COBOL Copybook Metadata?** – structure, definition

```
01  STUDENT.
  20  ID                              PIC 9(8).
  20  FIRST_NAME                      PIC X(32).
  20  LAST_NAME                       PIC X(32).
*                                     YYYYMMDD
  20  DATE_OF_BIRTH                   PIC S9(8) COMP.
  20  NUMOF_COURSES                   PIC 9(4) COMP.
  20  NUMOF_BOOKS                     PIC 9(4) COMP.
  20  COURSES.
    25 COURSE OCCURS 8 TIMES DEPENDING ON NUMOF_COURSES.
      30  COURSE_ID                   PIC 9(8).
      30  COURSE_TITLE                PIC X(48).
      30  INSTRUCTOR_ID               PIC 9(8).
      30  NUMOF_ASSIGNMENTS           PIC 9(4) COMP.
      30  ASSIGNMENTS OCCURS 4 TIMES DEPENDING ON NUMOF_ASSIGNMENTS.
        40  ASSIGNMENT_TYPE           PIC X(12).
        40  ASSIGNMENT_TITLE          PIC X(48).
*                                     YYYYMMDD
        40  DUE_DATE                  PIC S9(8) COMP.
        40  GRADE                     PIC S9V9.
  20  BOOKS.
    25  BOOK OCCURS 1 TO 5 TIMES DEPENDING ON NUMOF_BOOKS.
      30  ISBN                        PIC X(10).
*                                     YYYYMMDD
      30  RETURN_DATE                 PIC 9(8) COMP.
```

Metadata

Describes structure & format of data

The demand for COBOL & legacy metadata still strong, according to a recent DATAVERSITY survey.

Image Metadata

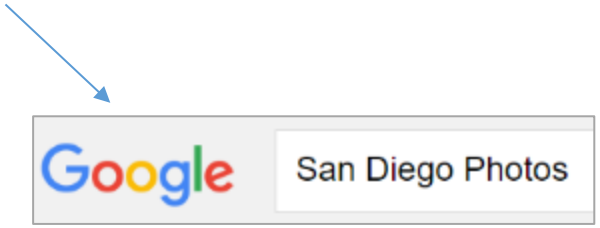
- Metadata is critical for locating images online, as well as identifying copyright information, etc.
- Some information is system-generated, while other is user-defined.

Technical Metadata
(Embedded in Photo)

Camera:	Apple iPhone 6 Plus
Lens:	iPhone 6 Plus back camera 4.15mm f/2.2 Shot at 4.2 mm Digital Zoom: 5.006134969×
Exposure:	Auto exposure, Program AE, 1/7,937 sec, f/2.2, ISO 32
Flash:	Auto, Did not fire
Date:	April 13, 2016 5:35:53PM (timezone not specified) (1 month, 11 days, 14 hours, 14 minutes, 46 seconds ago, assuming image timezone of US Pacific)
File:	3,264 × 2,448 JPEG (8.0 megapixels) 800,782 bytes (782 kilobytes)

Descriptive Metadata
(User Defined)

Title	DATAVERSITY EDW 2016 San Diego
Keywords	EDW 2016, San Diego, Bay Photos
Location	San Diego

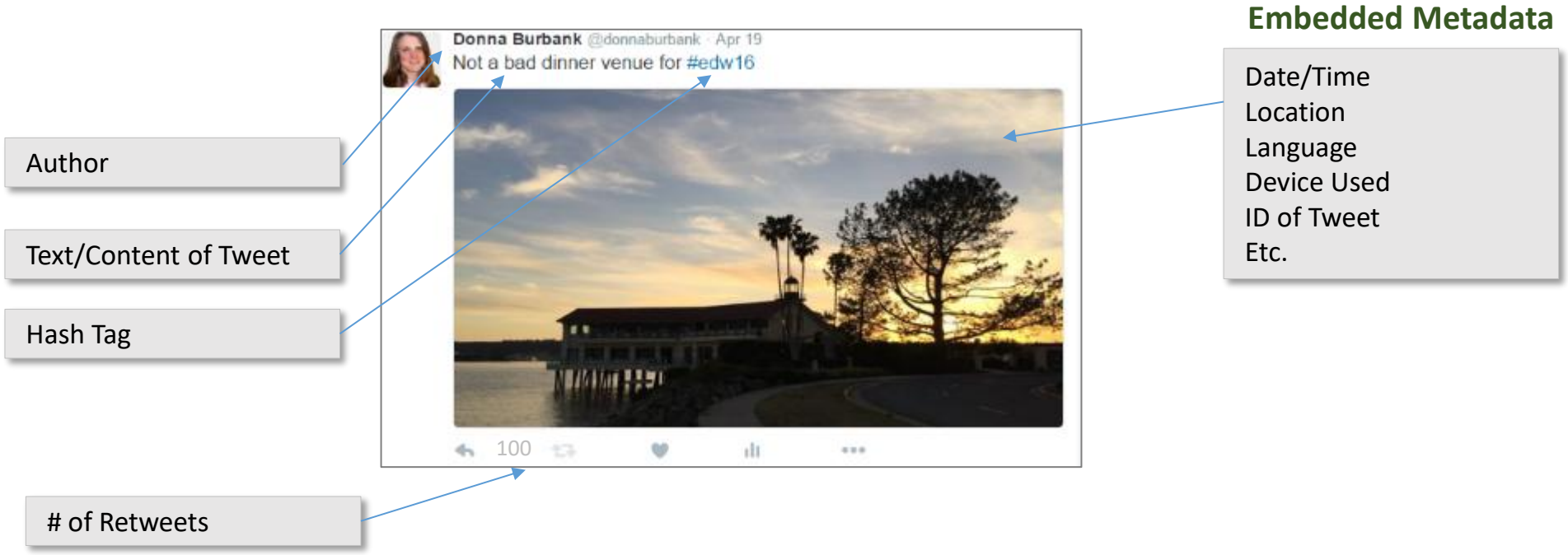


Administrative Metadata
(User Defined)

Author	Donna Burbank
Copyright	None
Licensing	None

Social Media Metadata

- Metadata from Social Media, such as X/Twitter, can help identify trend and sentiment analysis, for example.

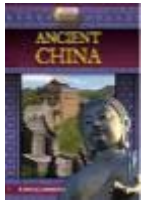


NoSQL Metadata – Document Databases

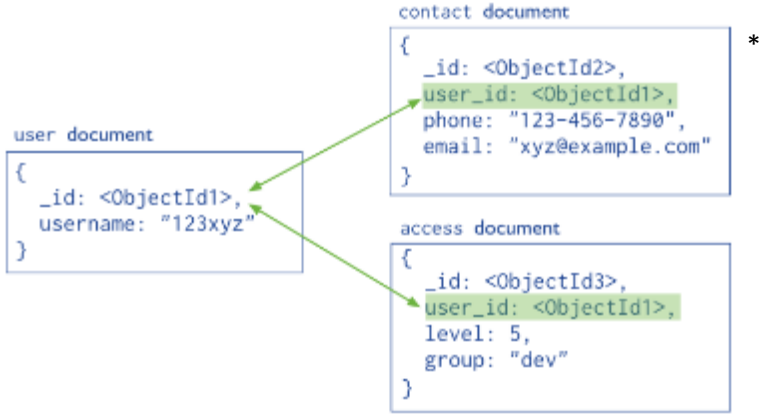
- Document databases are popular ways to store “unstructured” information in a flexible way (e.g. multimedia, social media posts, etc.)
- Each Collection can contain numerous Documents which could all contain different fields.

{type: “Artifact”,
medium: “Ceramic”
country: “China”,
}

{type: “Book”,
title: “Ancient China”
country: “China”,
}



- Some data modeling can be done, and some data modeling tools support this (e.g. MongoDB).



* Example from docs.mongodb.com

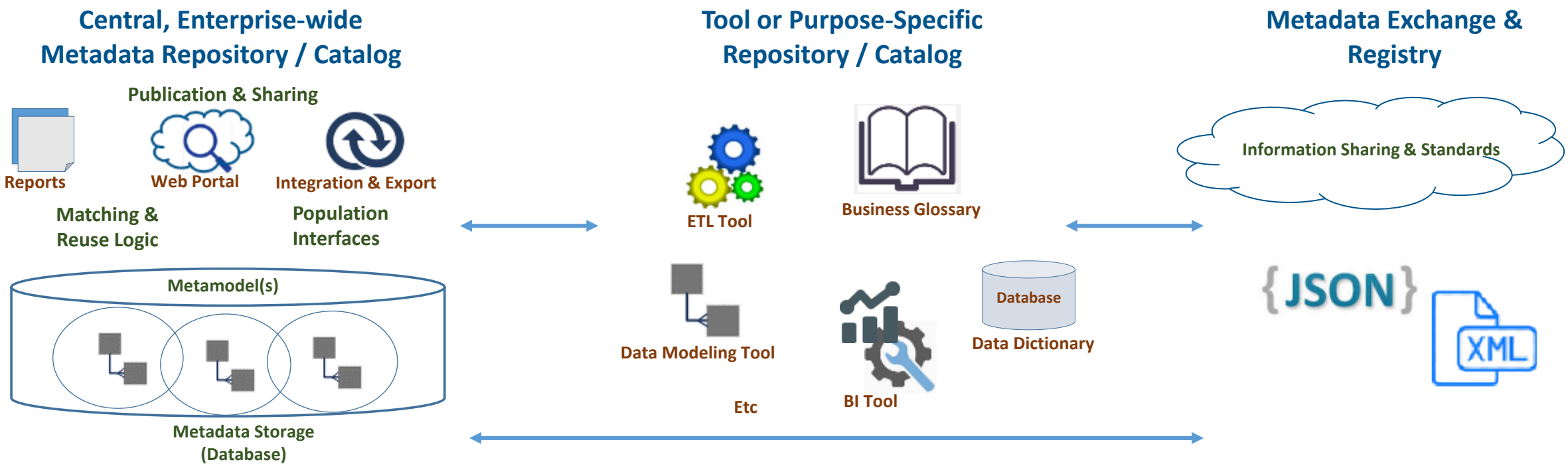
Vector Databases

- Vector Databases are used to represent images, audio text and other complex data (e.g. molecular structures).
- Images are converted into numerical vectors that capture **semantic meaning – i.e. metadata**
- Vector databases find data points that are close to each other in the vector space, based on similarity metrics.
- Video and Audio Search are common use cases for Vector Databases.

Find me a picture of a green frog on a bicycle.

Architectural Options for Metadata Management

- **The following are common architectural options for metadata management within & between organizations.**
  - There is no "one size fits all" approach.
  - They can be used together within the same organization.



Crowdsourcing Governance & Metadata Definitions

Many metadata projects & vendors are embracing the concept of "crowdsourcing".
i.e. The Wikipedia vs. Encyclopedia approach

Encyclopedia

- Created by a few, then published as read-only
- Single source of "vetted" truth
- Static

For Standardized, Enterprise Data Sets

Wikipedia

- Created by many, edited by many
- Eventual consistency with multiple inputs
- Dynamic

For Self-Service Data Prep & Analytics

WIKIPEDIA
The Free Encyclopedia

Finding the Right Balance

When implementing metadata management in today's rapidly-changing, self-service data landscape, it is important to find a balance between:

Standards-based Metadata & Governance

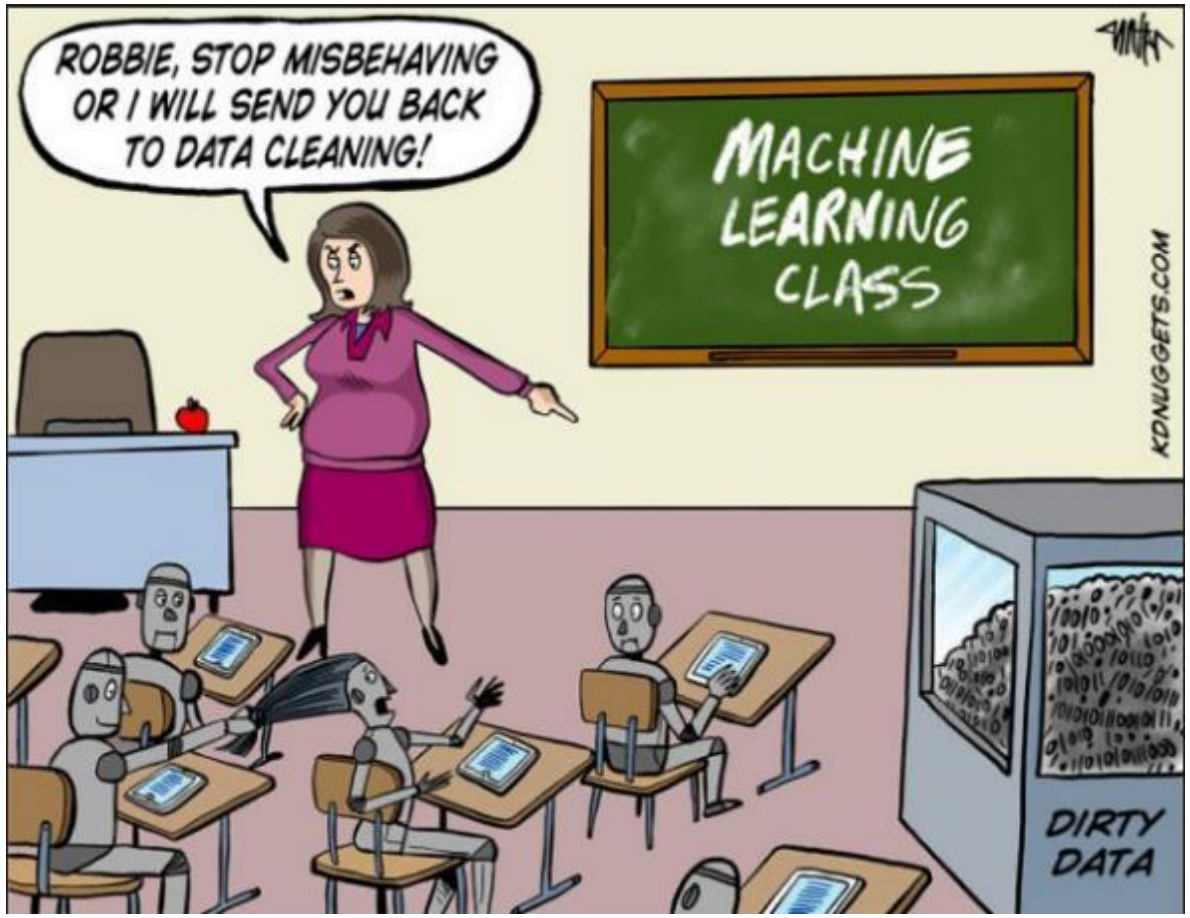
- Well-suited for enterprise-wide data standards

Collaboration-based Metadata & Governance

- Well-suited for self-service data preparation & analytics

The two methods work well together, using the right approach depending on the data usage.

Machine Learning & Metadata Discovery



Source kdnuggets.com

- Machine Learning offers ways to automate tedious tasks that may have been done manually before:
  - e.g. Data Mapping
    - SSN -> Field1_SSN
    - SSN -> Soc_Num
    - Etc.
  - Machine Learning Pattern Matching
    - NNN-NN-NNNN -> Field_X follows this pattern, it must be a SSN
- There is a place for both methods:
  - Sometimes you want to define specific mapping rules
  - Sometimes you want a pattern-matching, discovery-style approach.

Key Components of Metadata Management

Metadata Strategy	Metadata Capture & Storage	Metadata Integration & Publication	Metadata Management & Governance
Alignment with business goals & strategy	Identification of all internal & external metadata sources	Identification of all technical metadata sources	Metadata roles & responsibilities defined
Identification of & feedback from key stakeholders	Population/import mechanism for all identified sources	Identification of key stakeholders & audiences (internal & external)	Metadata standards created
Prioritization of key activities aligned with business needs & technical capabilities	Identification of existing metadata storage	Integration mechanism for key technologies (direct integration, export, etc.)	Metadata lifecycle management defined & implemented
Prioritization of key data elements/subject areas	Definition of enterprise metadata storage strategy	Publication mechanism for each audience	Metadata quality statistics defined & monitored
Communication Plan developed	Identification of business data stewards to populate business definitions	Feedback mechanism for each audience	Metadata integrated into operational activities & related data management projects

Summary

- **Metadata provides critical business and technical context** providing the "who, what, where, when, and why" around data
  - **Business metadata provides necessary context** around key data assets, and is often stored in the heads of key personnel
  - **Technical metadata can often be automated for metadata discovery**; human creation is typically necessary for design and creation
- **Metadata is relevant across a variety of use cases**, including Business Intelligence (BI) and Artificial Intelligence (AI)
- **A wide range of architectural options are available** for storing, sharing, and managing metadata within and between organizations.
- A successful metadata initiative should be **part of a wider data strategy**.

DATAVERSITY Data Architecture Strategies

This Year's Lineup

- **January** Trends in Data Architecture
- **February** Building a Data Strategy - Practical Steps for Aligning with Business Goals
- **March** Building the Right Architecture for Analytics & Reporting
- **April** Data Architect vs. Data Engineer vs. Data Scientist – Making Sense of Roles in Today's Data-Centric Organization
- **May** Master Data Management - Aligning Data, Process, and Governance
- **June** Where Data Models Fit in Today's Modern Data Architecture
- **July** Data Architecture vs. Enterprise Architecture
- **August** Data Quality Best Practices (with guest Nigel Turner)
- **September** Modern Data Architecture: Practical Options for Today's Data-Driven Organization
- **October** Best Practices in Metadata Management
- **December** The Business Value of Data Modeling

Who We Are: Business-Focused Data Strategy

Maximize the Organizational Value of Your Data Investment

In today's business environment, showing **rapid time to value** for any technical investment is critical.

But technology and data can be complex. At Global Data Strategy, **we help demystify technical complexity** to help you:

- Demonstrate the ROI and **business value of data**
- Build a data strategy **at your pace to match your unique culture** and organizational style.
- Create an **actionable roadmap for "quick wins"**, which building towards a long-term scalable architecture.

Global Data Strategy shares experience from some of the largest international organizations scaled to the pace of your unique team.

Global Data Strategy has worked with organizations globally in the following industries:

Finance · Retail · Social Services · Health Care · Education · Manufacturing · Government · Public Utilities · Construction · Media & Entertainment · Transportation · Insurance …. and more

Thoughts? Ideas?

Questions?