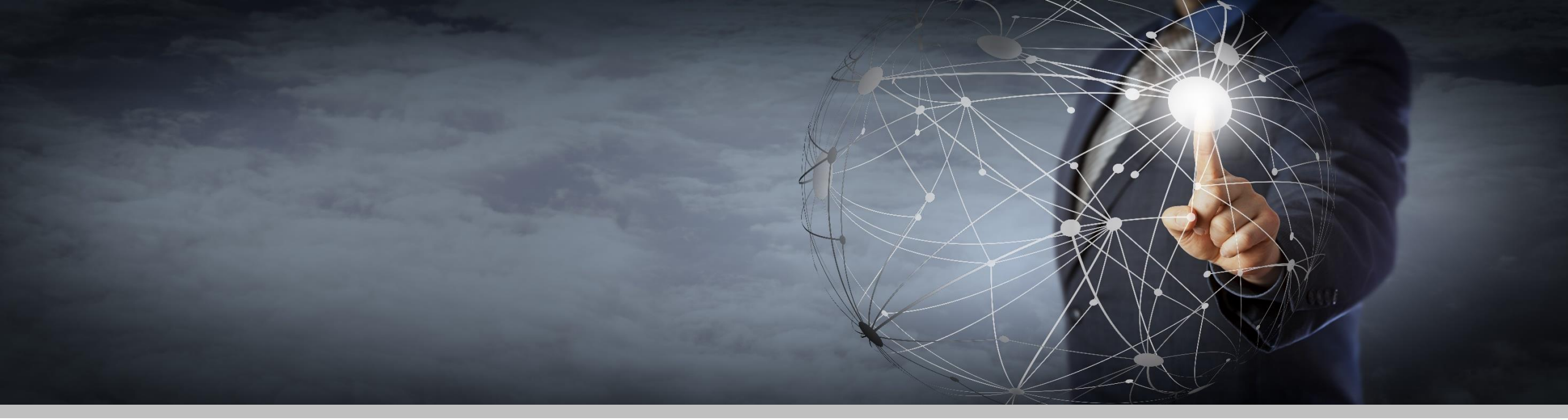


Best Practices in Metadata Management

Donna Burbank
Global Data Strategy, Ltd.

Sept 26, 2024

DATA ARCHITECTURE STRATEGIES

Global Data Strategy
The Business of Data

Donna Burbank

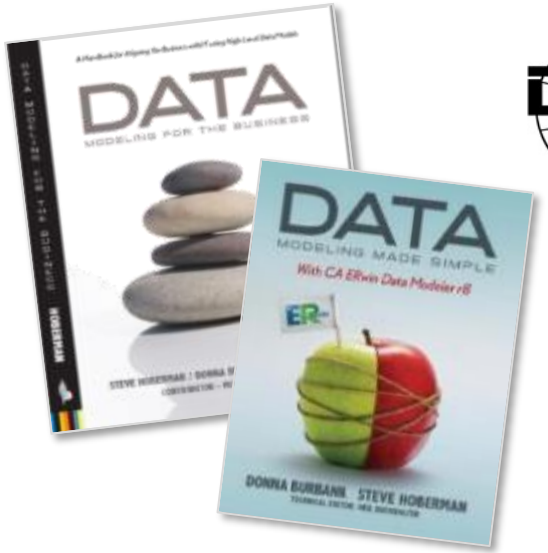
Donna is a recognized industry expert in data management with over 25 years of experience in data strategy, data governance, data modeling, metadata management, and enterprise architecture. Her background is multi-faceted across consulting, product development, product management, brand strategy, marketing, 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 International member, contributor to the DMBOK 2.0, 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.



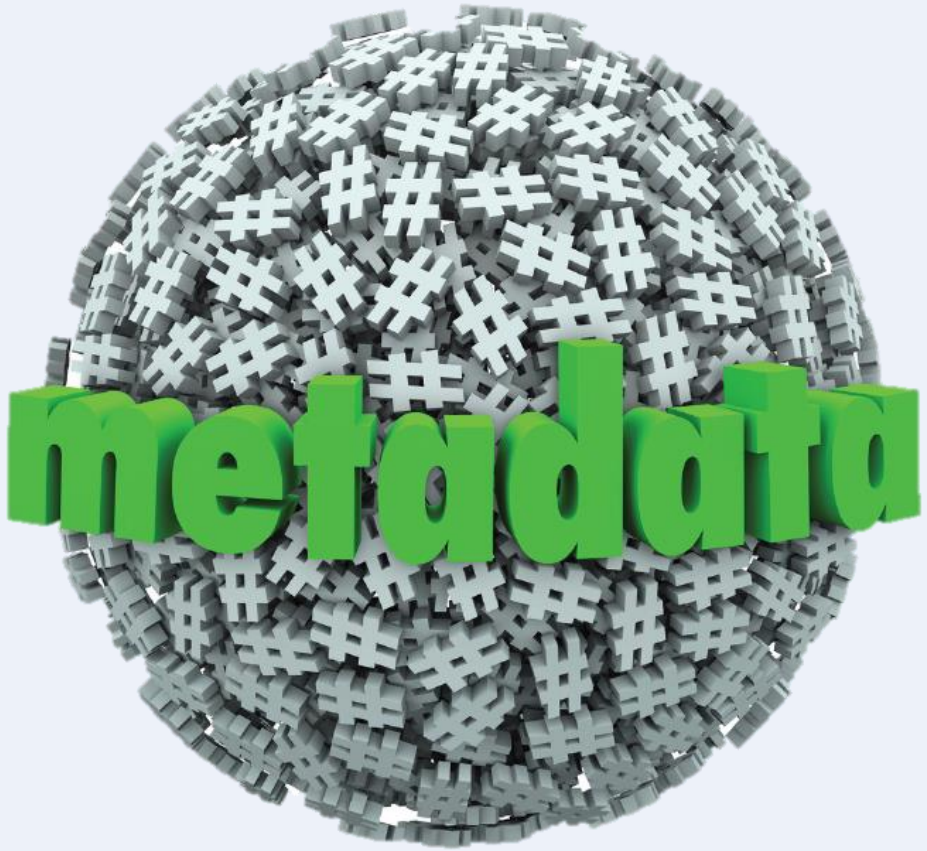
DATAVERSITY Data Architecture Strategies

This Year's Lineup

- **January** Emerging Trends in Data Architecture – What's the Next Big Thing?
- **February** Building a Data Strategy - Practical Steps for Aligning with Business Goals
- **March** Master Data Management - Aligning Data, Process, and Governance
- **April** How do Data Governance & Data Architecture Support Each Other?
- **May** The Role of the Chief Data Officer (CDO) in Business Transformation
- **June** What Does It Mean to be a Data-Driven Organization?
- **July** Data Architect vs. Data Engineer vs. Data Scientist – Making Sense of Roles in Today's Data-Centric Organization
- **August** Data Quality Best Practices (with Nigel Turner)
- **September** Best Practices in Metadata Management
- **October** Enterprise Architecture vs. Data Architecture
- **December** The Business Benefits of Data Modeling

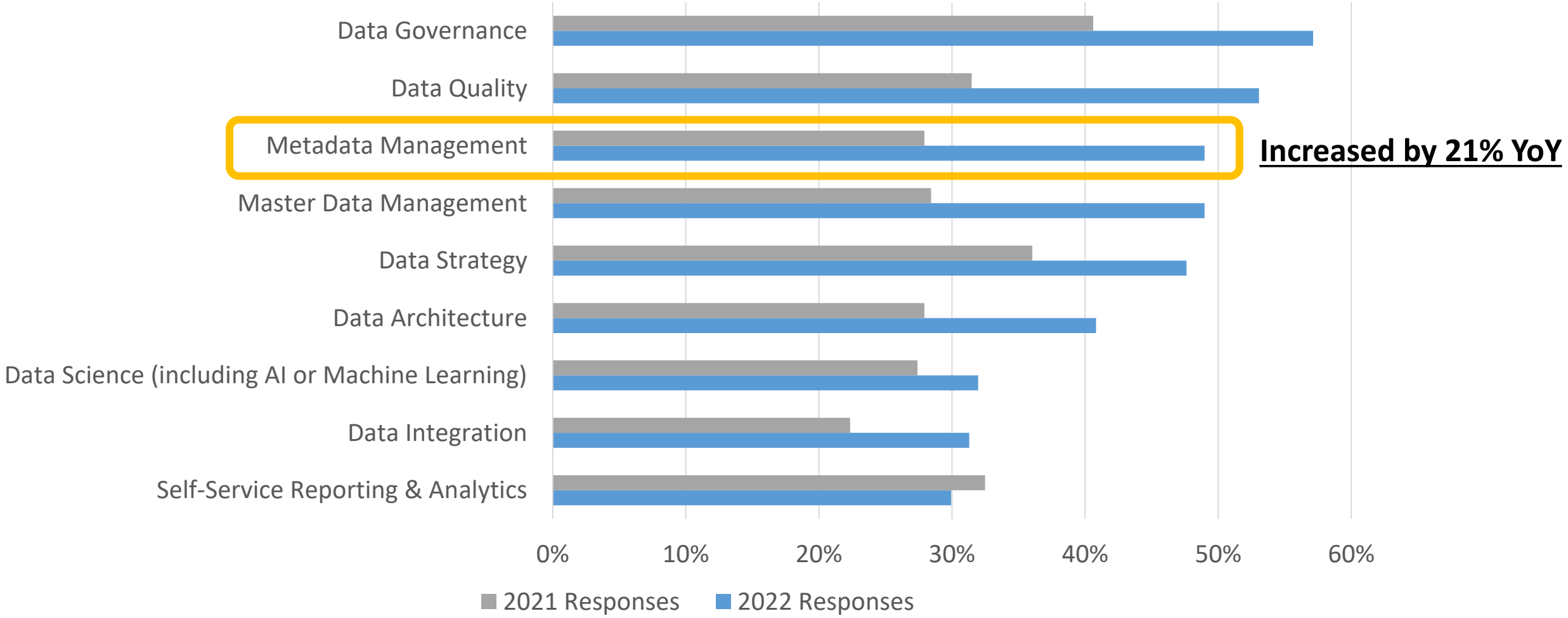
What We'll Cover Today

- **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.
- At the same time, **industry regulations are driving the need for better transparency** and understanding of information.
- 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 Initiatives for Organizations in 2023-2024

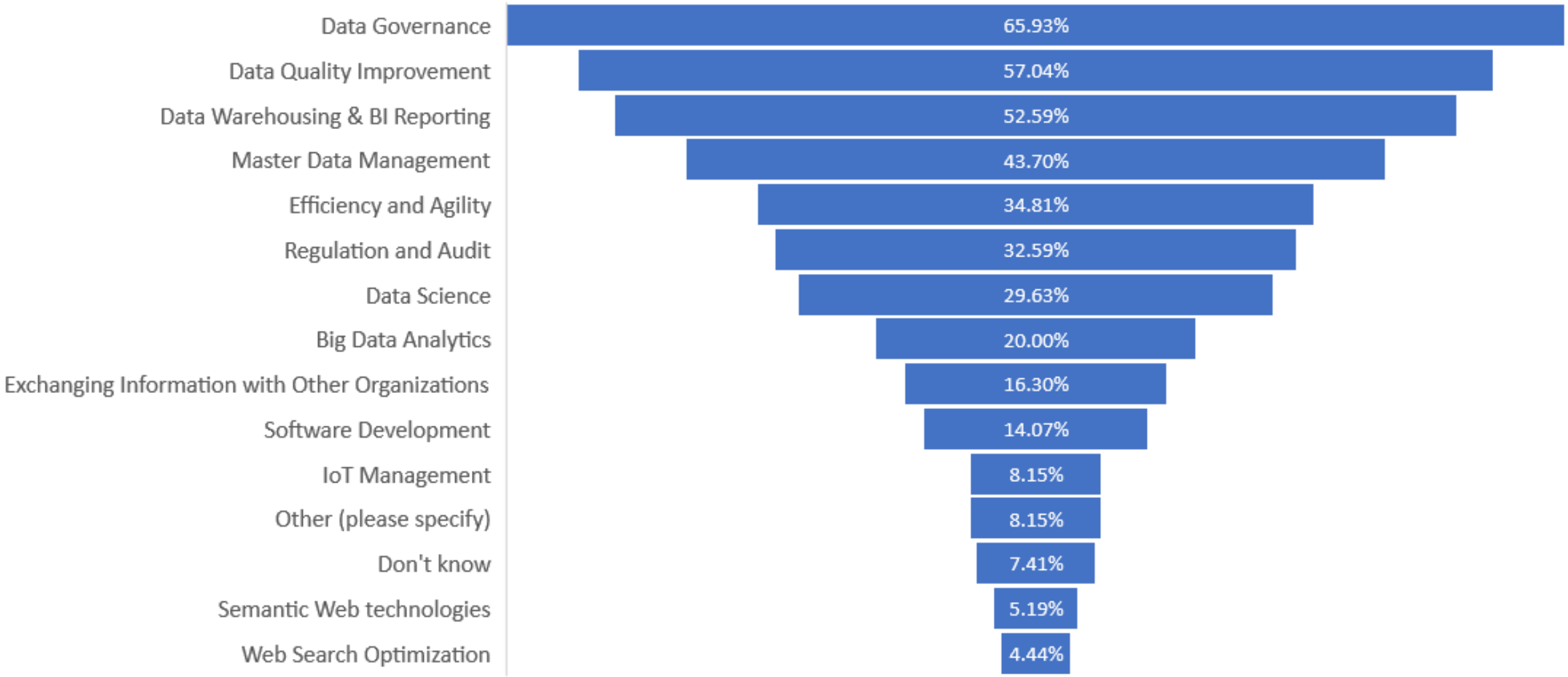
Metadata is a Top Priority for Organizations in Coming Years



Increased by 21% YoY

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

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

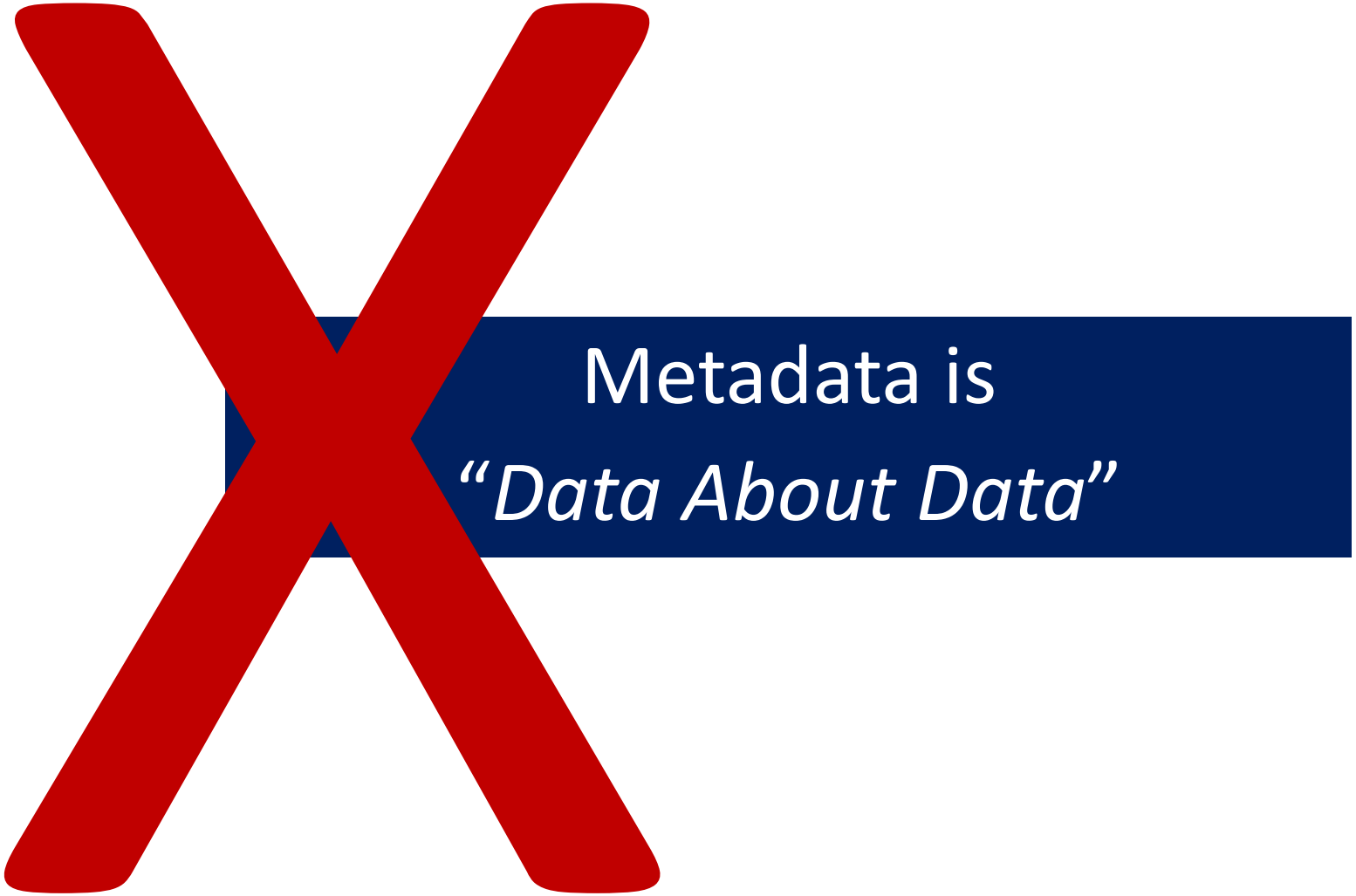
Managing the people, process, policies & culture around data

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?

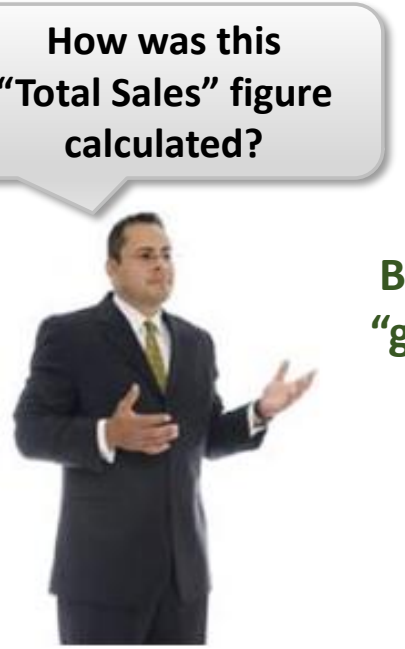
Definition	Last Name represents the surname or family name of an individual.
Business Rules	In the Chinese market, family name is listed first in salutations.
Format	VARCHAR(30)
Abbreviation	LNAME
Required	YES
Etc.	Numerous technical & business metadata including security, privacy, nullability, primary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

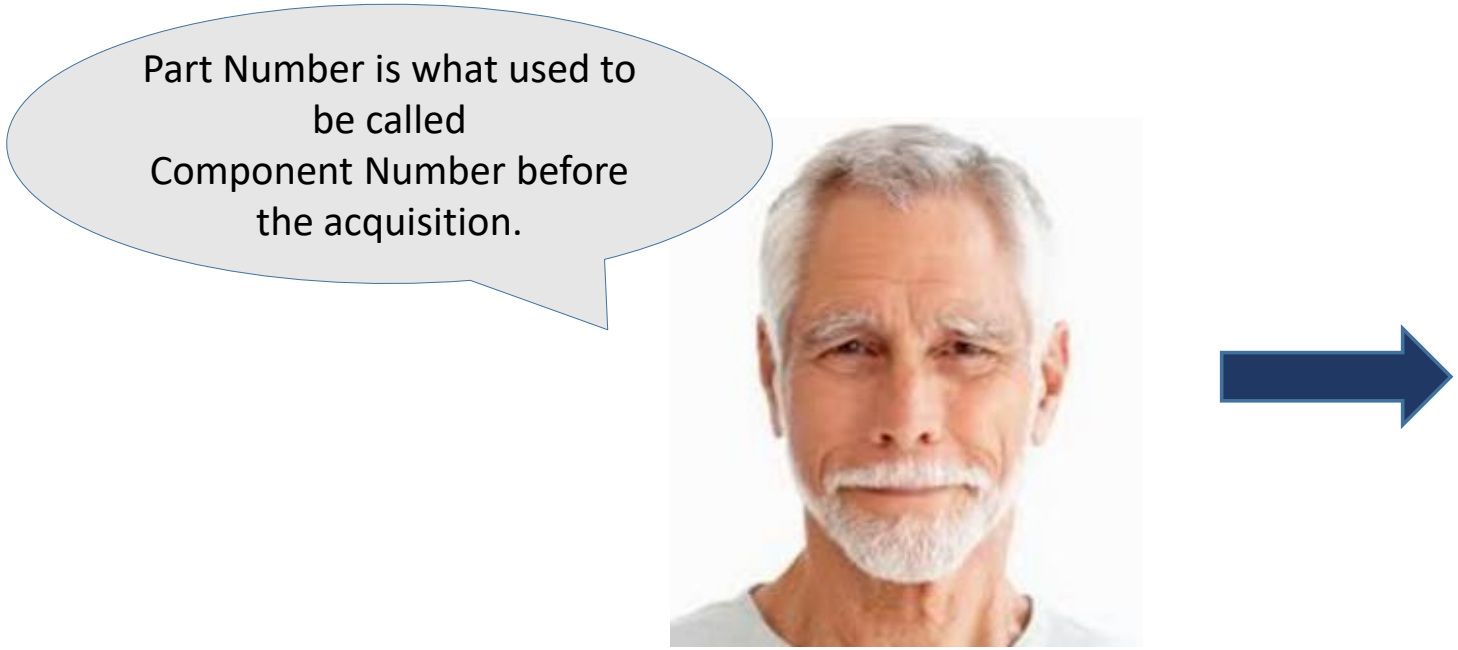
"Do we have to obfuscate PII?"

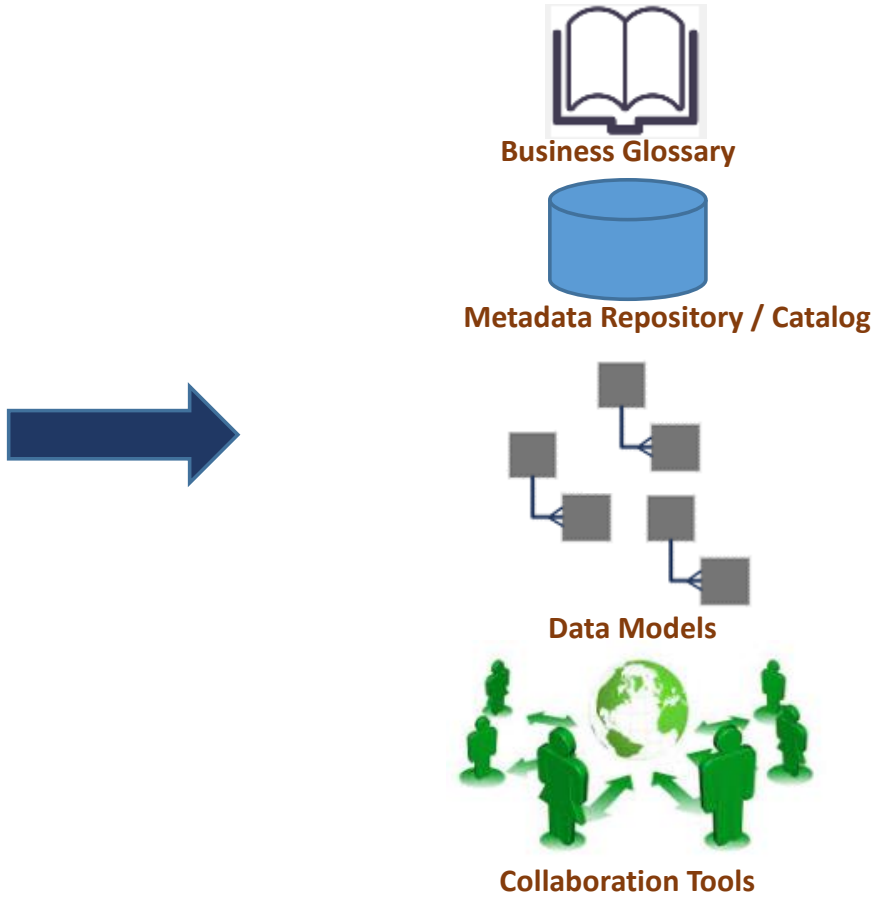
Data Architect

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”





Business Definitions

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

Better Definitions Drive Better Communication

- Wouldn't it be helpful if we did this in daily life, too?
- i.e. "Let's go on a family vacation!"

Person	Concept	Definition
Father	Vacation	An opportunity to take the time to achieve new goals
Mother	Vacation	Time to relax and read a book
Jane	Vacation	A chance to get outside and exercise
Bobby	Vacation	Time to be with friends
Cousin Ian	Holiday	An excuse to go to the pub
Donna	Vacation	More time to design metadata architectures

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

NASA Open Data (with Metadata)

GODDARD SPACE FLIGHT CENTER
Space Physics Data Facility

+ Goddard Home
+ NASA Home

Space Physics Data Facility

+ ABOUT | + DATA & ORBITS | + ModelWeb at CCMC | + SCIENCE ENABLED | + AND MORE

Data Access & Orbit Services

+ Heliophysics Data (search) Portal
+ Gateway to SPDF Services
+ CDAWeb (data browser)
+ CDAWeb Inside IDL
+ OMNIWeb Plus (now including COHOWeb, ATMOWeb, FTP Browser, HelioWeb and CGM)
+ Direct HTTP(S) to Data
+ Direct FTP(S) to Data (FTPS required)
+ SSCWeb (orbit search)
+ 4D Orbit Viewer
+ GIFWalk data and orbit plots
+ Alternative Data Access Methods
+ SDAC VSO - Virtual Solar Observatory
+ SDAC - Solar Data Analysis Center
+ More information on Data Access for New Users

NASA's Space Physics Data Facility (SPDF)

Space Physics Data Facility (SPDF) is the NASA active and permanent archive for non-solar heliophysics data (solar data at SDAC), per the NASA Heliophysics Science Data Management Policy. SPDF is a project of the Heliophysics Science Division (HSD) at NASA's Goddard Space Flight Center. SPDF also provides multi-project, cross-disciplinary access to data to enable correlative and collaborative research across discipline and mission boundaries with present and past missions. SPDF maintains the SSCweb database of spacecraft orbits, the OMNIweb cross-normalized database, and the Common Data Format (CDF) self-describing science data format and associated software.

News & Announcements

NOTICE: July 2021: The Parker Solar Probe (PSP) data have been extended to March 2021, which includes Encounter 7, the rest of Orbit 7, and the 4th Venus flyby. Some SWEAP SPAN data sets had new variables added. The Fluxgate magnetic field data are reprocessed for the entire mission. The merged fluxgate and search coil magnetic field data are updated for Encounters 1-3, and the high-rate EPI-Hi data of ISOIS from 2020-11-30 to 2020-12-02 are not fully calibrated yet.

NOTICE: April 2021: Global-scale Observations of the Limb

SPDF Web Service APIs

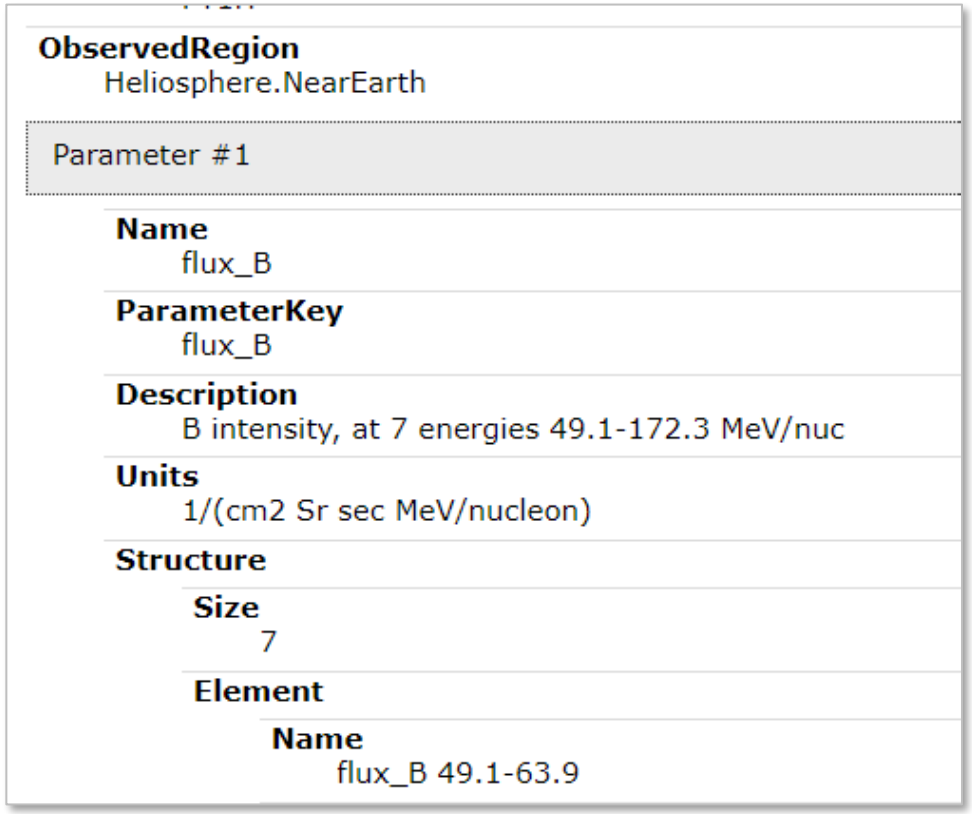
+ CDAWeb
+ SSCWeb
+ Heliophysics API (HAPI)

Software

+ CDF (Common Data Format)
+ Space Physics use of CDF
+ CDF/netCDF/FITS/ HDF/XML/ASCII Format Translations
+ CDF SKTEditor
+ MakeCDF
+ CDAWlib /CDFX (IDL)
+ ViSBARD (visualization)

Submit New Data to the Archive

+ New mission data requirements
+ Overview of SPDF Data Submission Guidelines and Procedures



ObservedRegion
Heliosphere.NearEarth

Parameter #1

Name
flux_B

ParameterKey
flux_B

Description
B intensity, at 7 energies 49.1-172.3 MeV/nuc

Units
1/(cm2 Sr sec MeV/nucleon)

Structure

Size
7

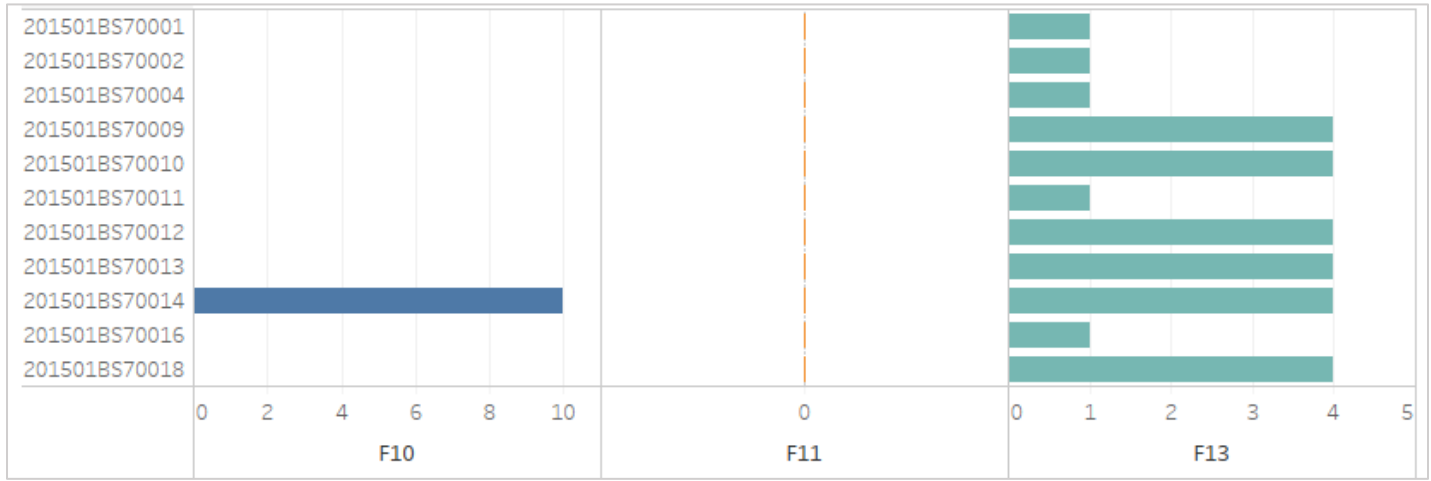
Element

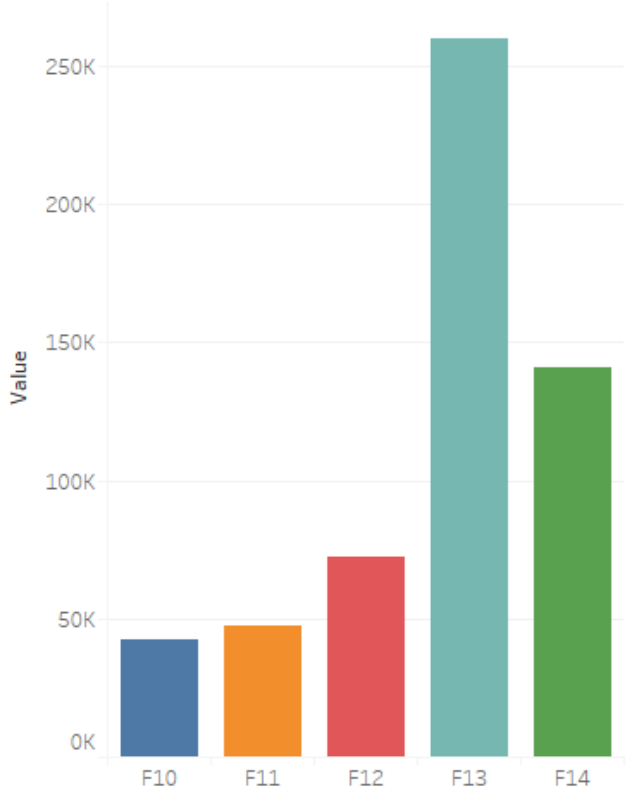
Name
flux_B 49.1-63.9

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

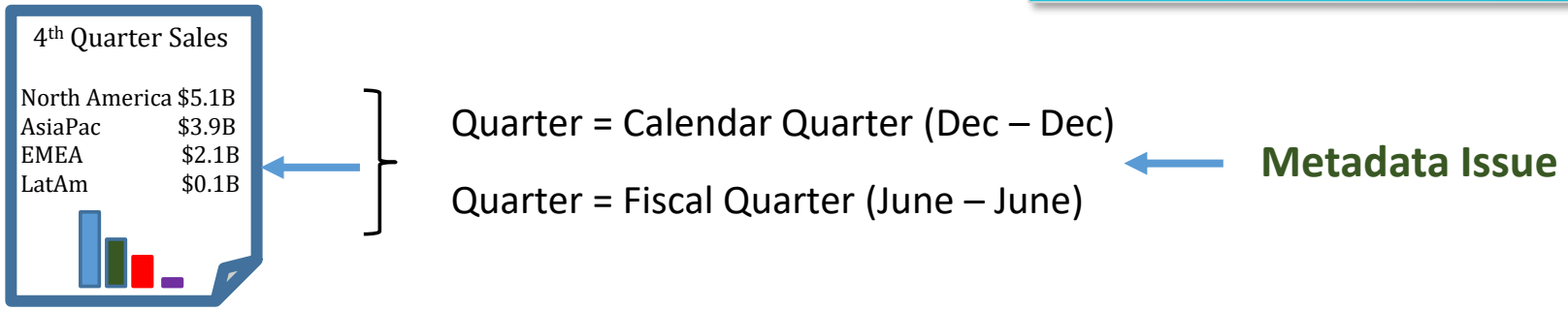




Financial Reporting – What is a Year?

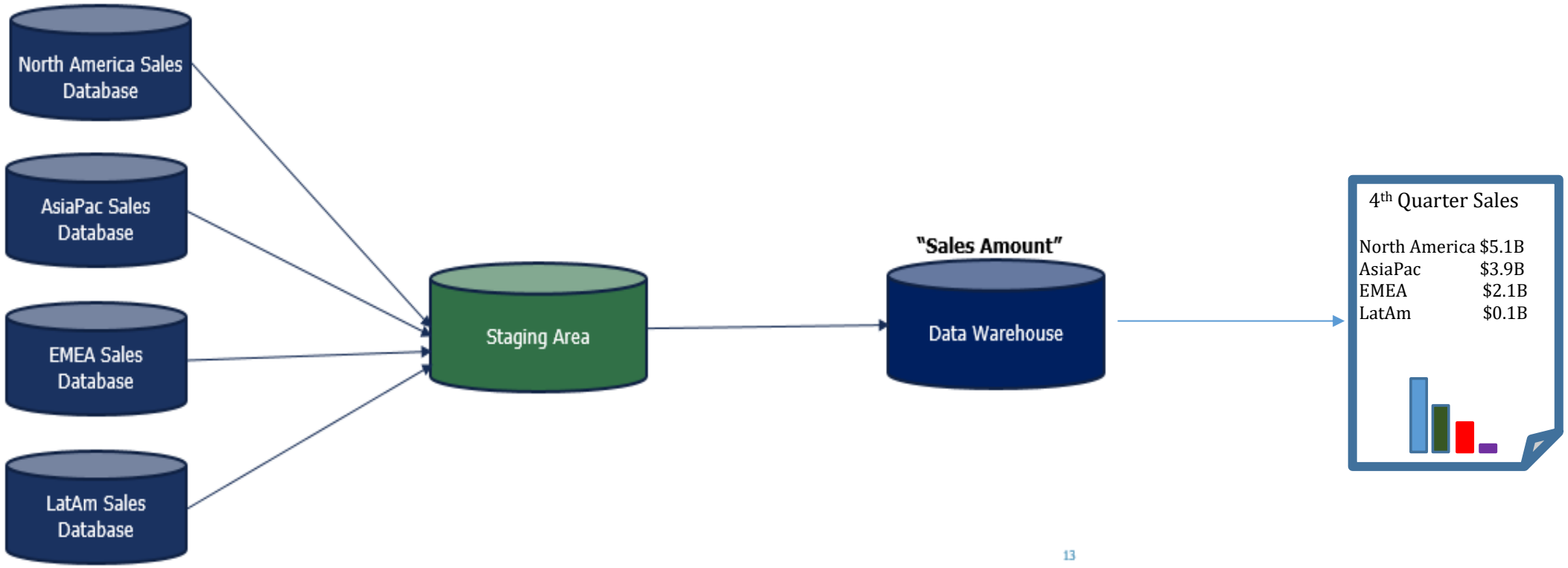
An international retail chain was comparing
4th Quarter Sales across regions.

- **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.**



Who Uses Metadata?

Data Governance is a Critical Enabler for Metadata Management

- Data Governance creates the roles, policies, procedures, and organizational structures to facilitate metadata management.
- Multiple Roles work together to create business and technical metadata.

Sample Governance Roles Involved with Metadata Creation *

Business Data Owner

- KPIs
- Organizational Metrics
- Regulatory Guidelines & Policy

Business Data Steward

- Glossary terms & definitions
- Business rules
- Acronyms

Data Architect

- Conceptual & logical models w/ core business rules and definitions
- Naming standards
- Data Lineage

System Data Steward

- Physical metadata structures for core applications
- Business definitions for application fields
- Alignment of systems with business rules

DBA/Data Engineer

- Physical metadata structures
- Naming standards
- Data type standards

Policies, Procedures, Training, and Job Descriptions help guide and enforce metadata creation and maintenance.

* Note: Roles are different for each organization. Each organization's governance structure and roles are unique.

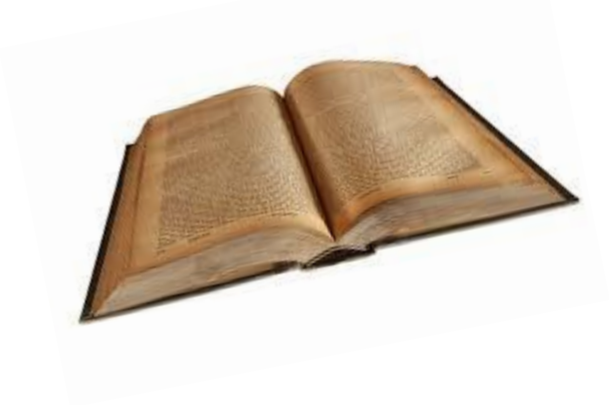
Crowdsourcing Governance & Metadata Definitions

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

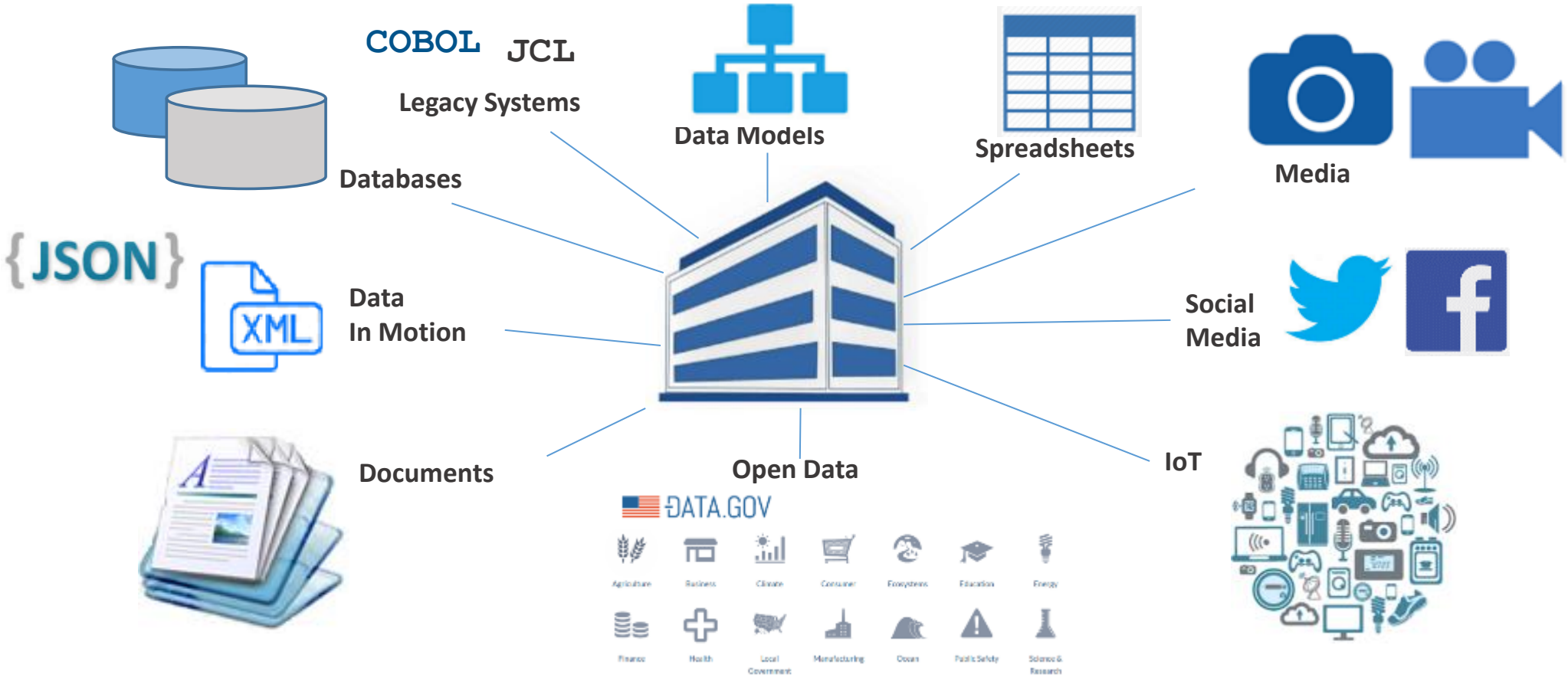
Collaboration-based Metadata & Governance

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

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

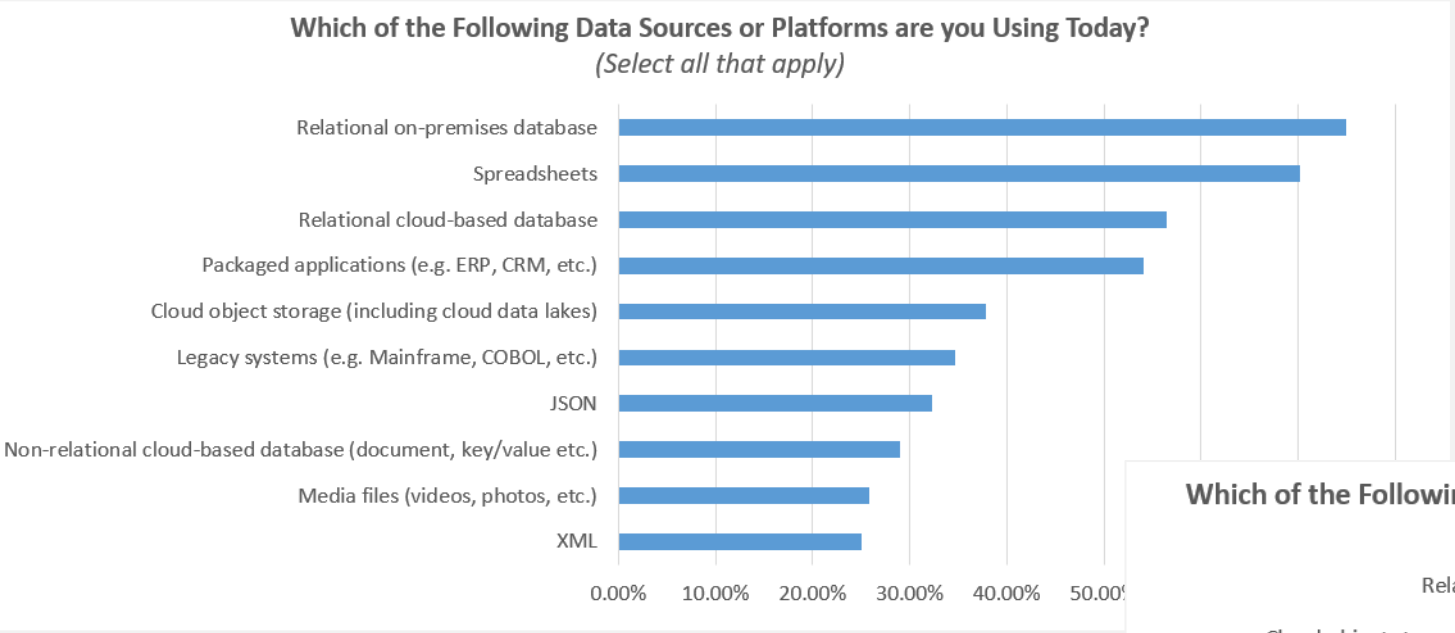
Metadata Across & Beyond the Organization

- Metadata exists in many sources across & beyond the organization.



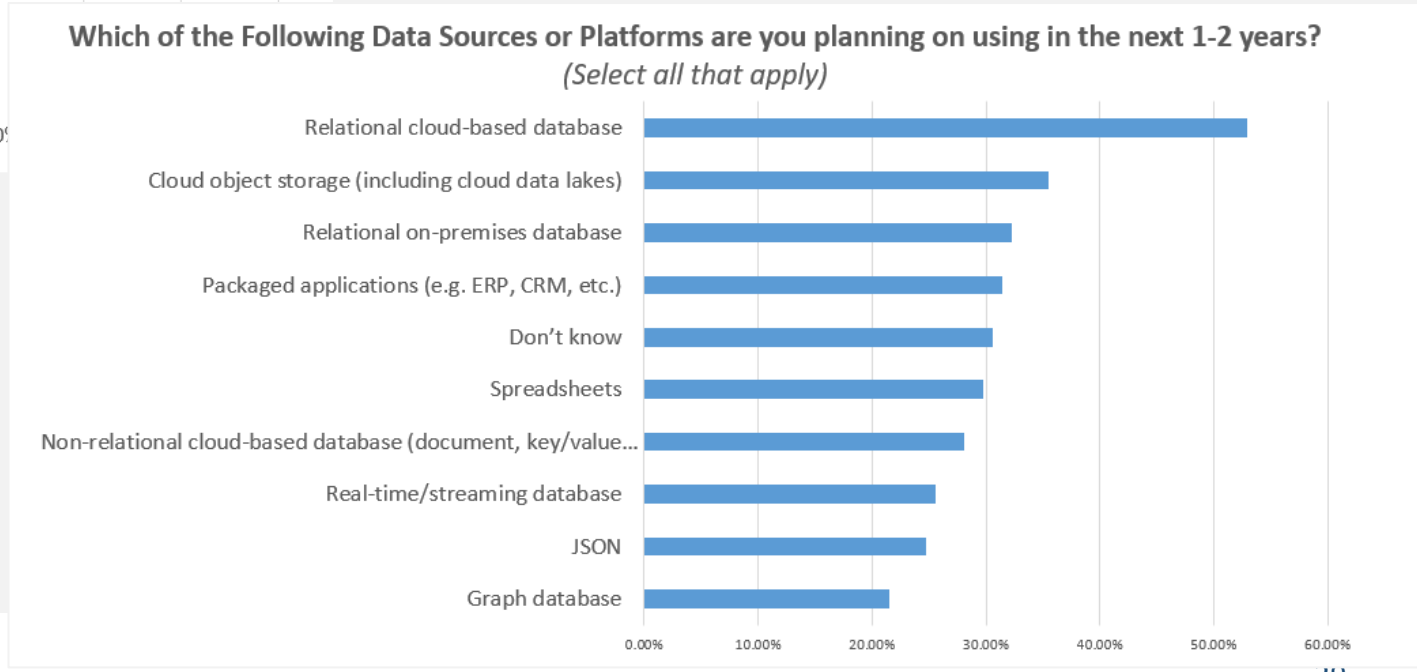
Metadata Sources

Current



The DATAVERSITY *Trends in Data Management* survey revealed some interesting findings about what types of data platforms (metadata sources) organizations will be managing now and in the future.

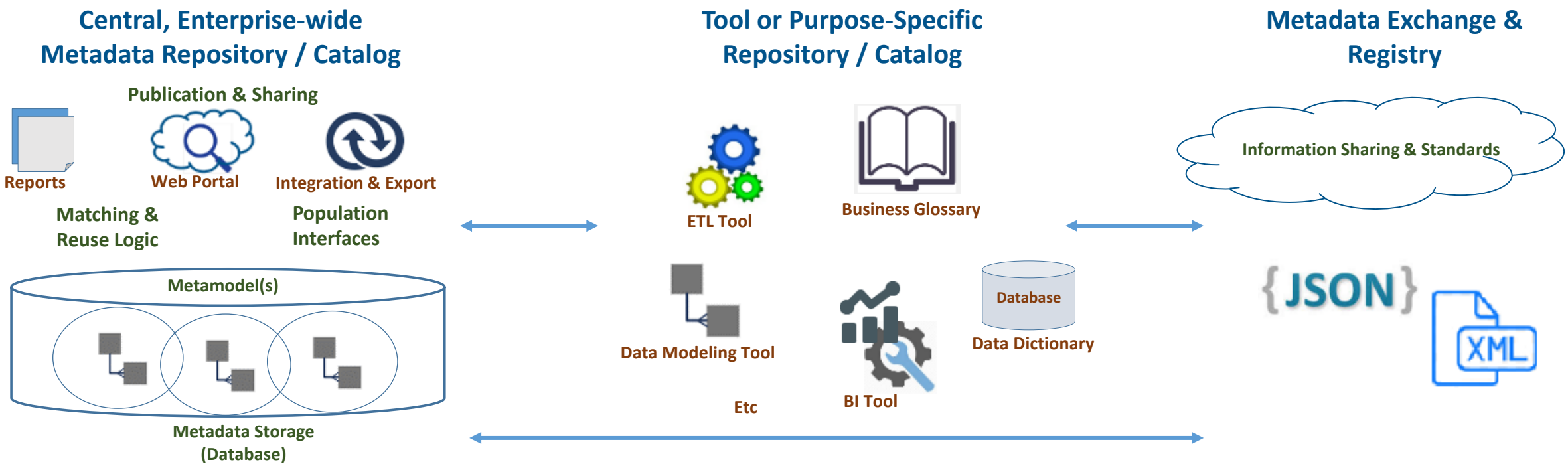
Future



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

Architectural Options for Metadata Management

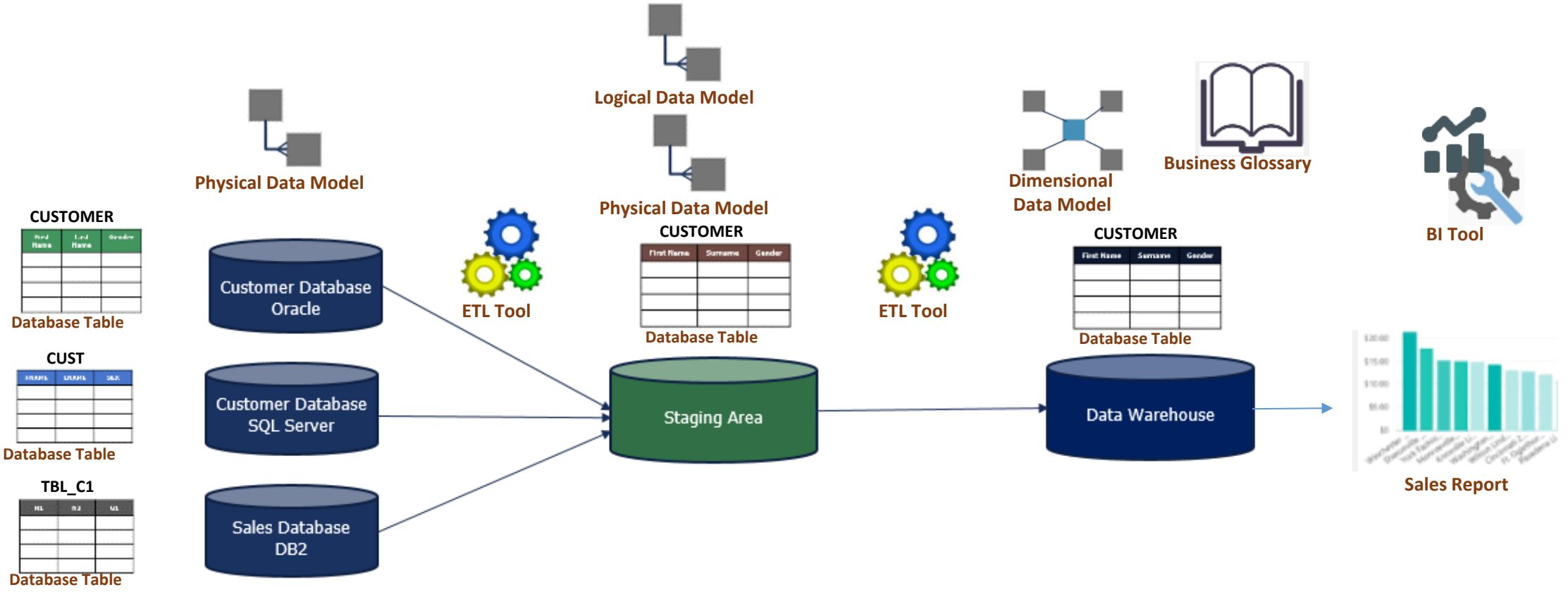
- **The following are common architectural options for metadata management within & between organizations.**
  - There is no "one size fits all" approach.
  - They can be used together within the same organization.



Data Lineage

Data Warehousing Example

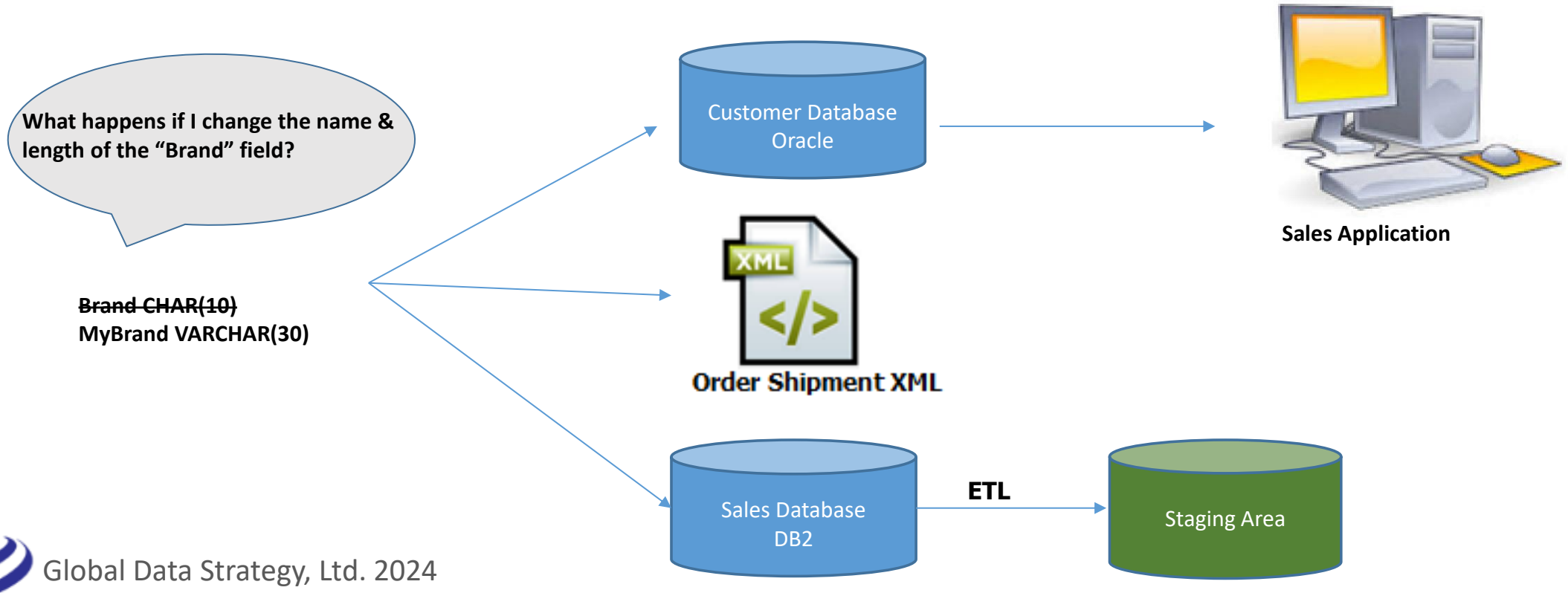
- In the data warehouse example below, metadata for CUSTOMER exists in a number tools & data stores.



Impact Analysis & Where Used

Showing the Impact of Change

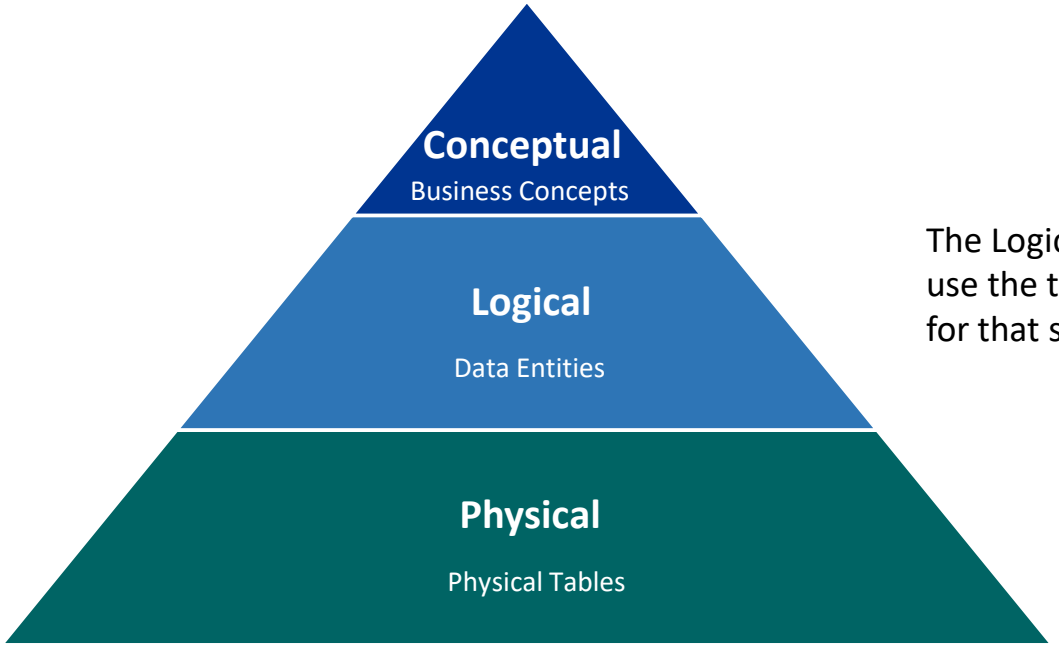
- **Impact Analysis** shows the relationship between a piece of metadata and other sources that rely on that metadata to assess the impact of a potential change.
- For example, if I change the length & name of a field, what other systems that are referencing that field will be affected?

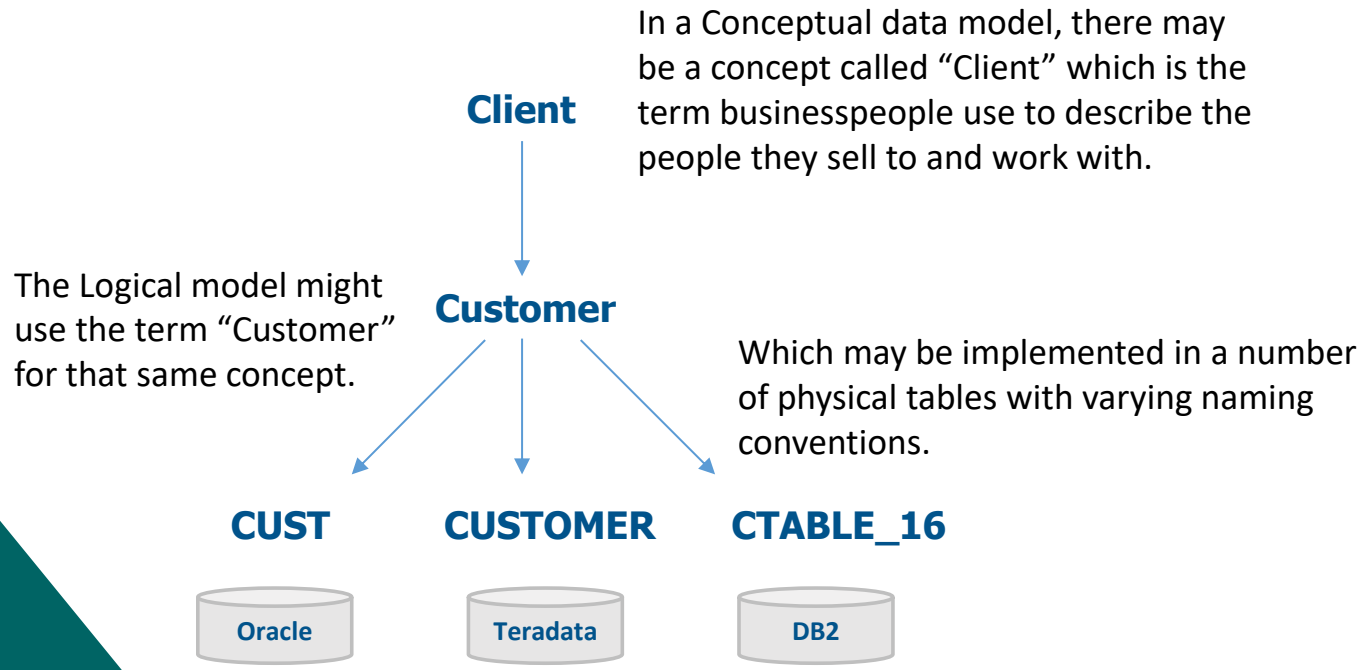


Semantic or Design Layer Relationships

Showing Semantic Mapping

- For example, data model design layer mappings show the relationship between business terms and their physical implementations on a database platform.
- Many metadata repositories have similar business-to-technical mapping & lineage.

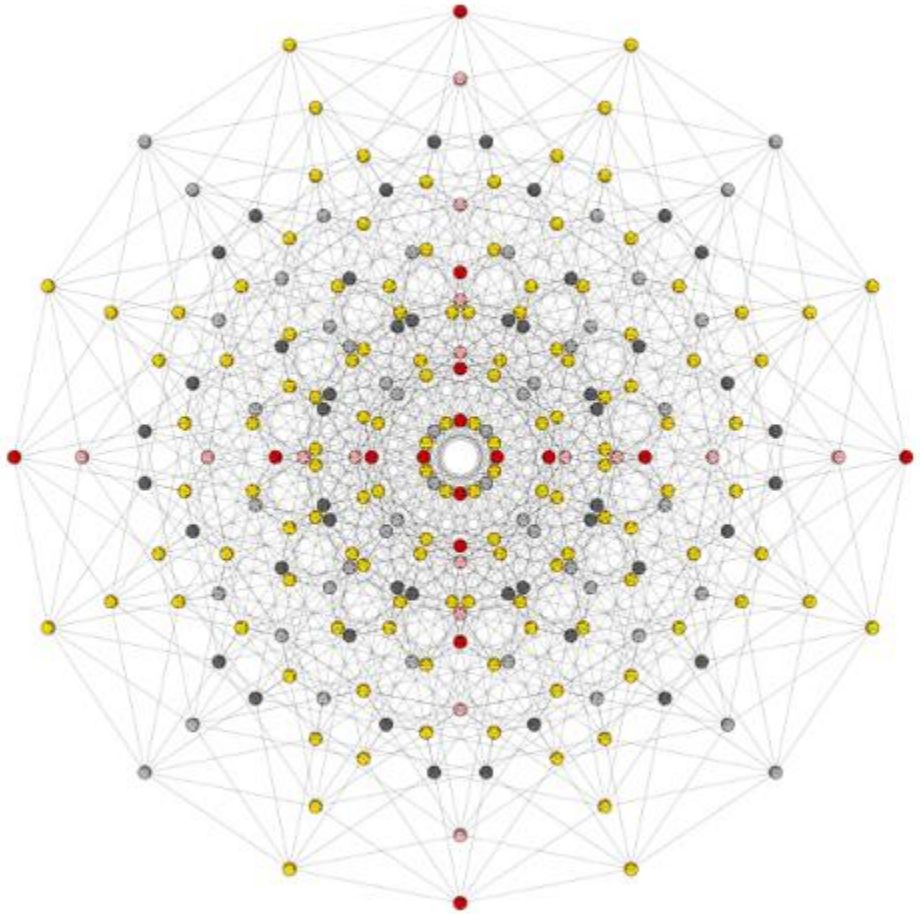




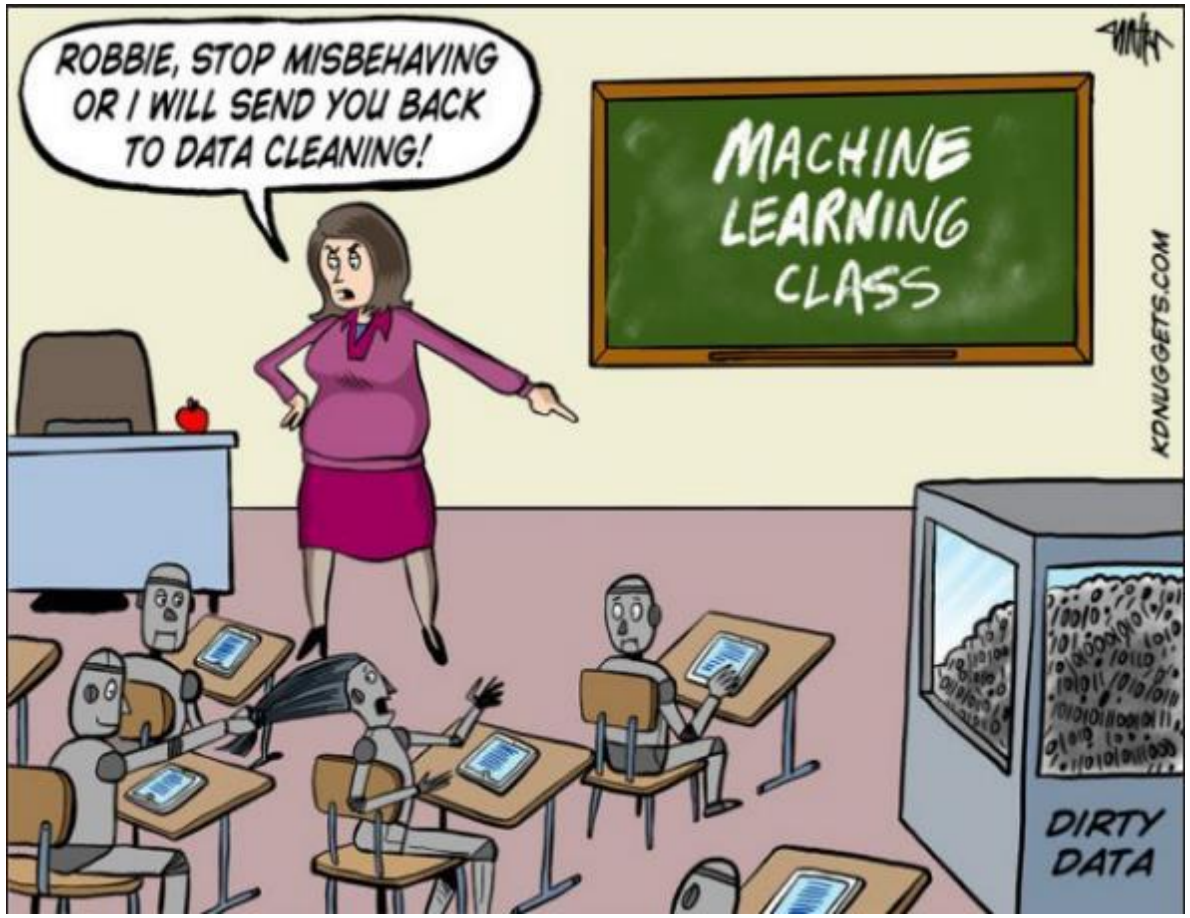
Graph Relationships

Patterns & Interrelationships

- Graph databases are ideal for analyzing metadata relationships between objects and finding patterns in those relationships.
- Common use cases for graph relationship metadata analysis include:
  - **Fraud detection** - e.g. financial transactions
  - **Threat detection** - e.g. email and phone patterns
  - **Marketing** – e.g. social media connections, product recommendation engines
  - **Network optimization** - e.g. IoT, Telecommunications



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
- **Data governance provides orchestration** for roles and responsibilities around metadata creation and maintenance
  - **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
- **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** Emerging Trends in Data Architecture – What's the Next Big Thing?
- **February** Building a Data Strategy - Practical Steps for Aligning with Business Goals
- **March** Master Data Management - Aligning Data, Process, and Governance
- **April** How do Data Governance & Data Architecture Support Each Other?
- **May** The Role of the Chief Data Officer (CDO) in Business Transformation
- **June** What Does It Mean to be a Data-Driven Organization?
- **July** Data Architect vs. Data Engineer vs. Data Scientist – Making Sense of Roles in Today's Data-Centric Organization
- **August** Data Quality Best Practices (with Nigel Turner)
- **September** Best Practices in Metadata Management
- **October** Enterprise Architecture vs. Data Architecture
- **December** The Business Benefits 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 · Insurance …. and more

Thoughts? Ideas?

Questions?