

Data Integration Demystified: Enabling AI-Driven Interoperability

Using Metadata, Cloud, and Warehousing Examples



peter.aiken@anythingawesome.com +1.804.382.6957

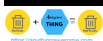


© Copyright 2026 Peter Aiken, PhD Slide # 1

Peter Aiken, Ph.D.

- I've been doing this a long time
- My work is recognized as useful
- Associate Professor of IS (vcu.edu)
- DAMA International (dama.org)
 - IDEAL Affinity Group
 - GenAI in Data Affinity Group
- Anything Awesome (anythingawesome.com)
- Experienced w/ 500+ data management practices worldwide
- 13 books and dozens of articles
- Multi-year immersions
 - US DoD (DISA/Army/Marines/DLA)
 - Nokia
 - Deutsche Bank
 - Wells Fargo
 - Walmart
 - HUD
 - NGA ...

\$1,500,000,000.00 USD



anythingawesome.com



Slide # 2

Program Overview

*Data Integration Demystified:
Enabling AI-Driven Interoperability*

- Integration Goals/Challenges
- AI-based approach-augmentation
- Metadata
 - Replace the concept of technical detail with that of a foundational language (common vocabulary) that provides context and clarity for your data and AI
- Cloud
 - Before taking full advantage of the scalability, agility, and cost-effectiveness of cloud services, cloud data should be architected
- Data Warehousing
 - Needs a new layer to prepare for AI
- Take aways, references and Q&A



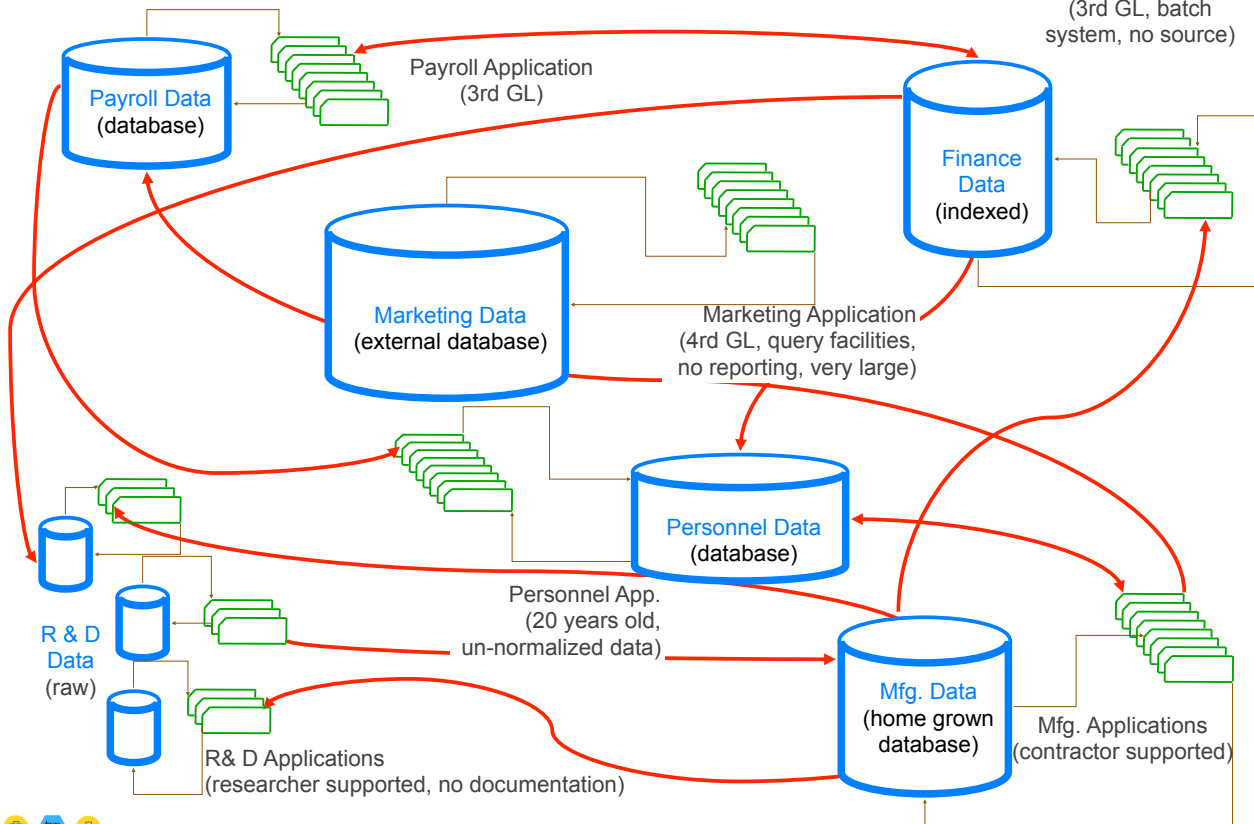
Integration Goals

- Improving the overall performance of large, complex systems in areas such as:
 - Processing efficiency
 - Unit responsiveness
 - Perceived quality
 - Product differentiation.
- Integration efforts seek to achieve improvements by facilitating the interaction among:
 - Organizations
 - Individuals, and
 - Systems

Goals

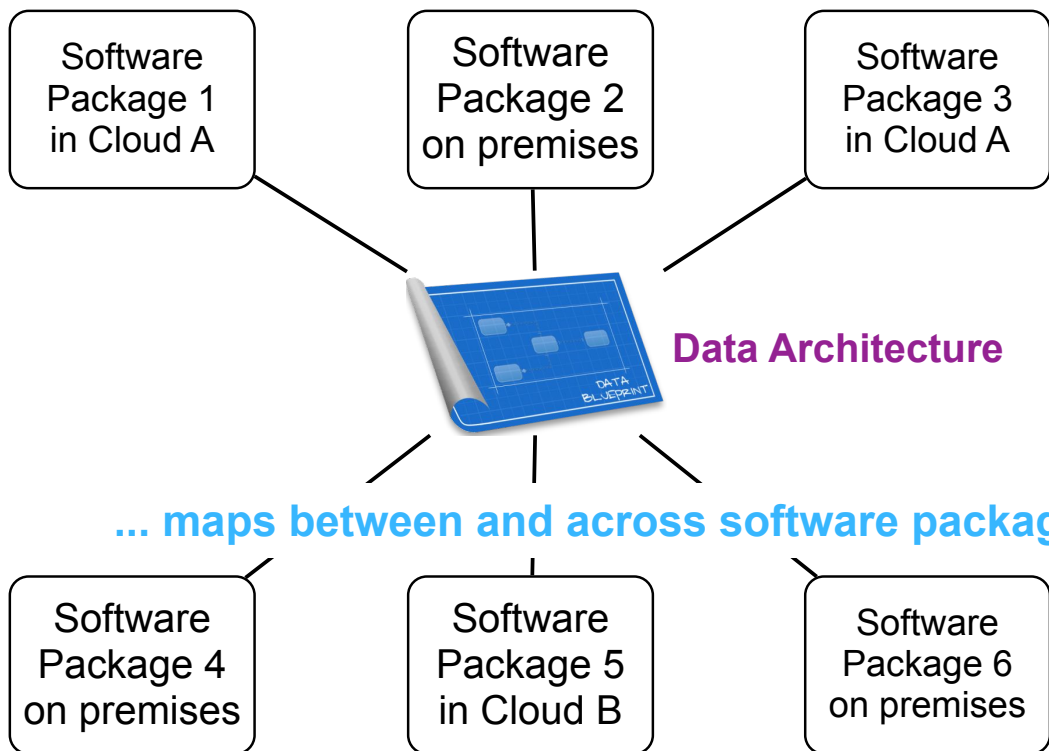


Typical Organizational System Evolution



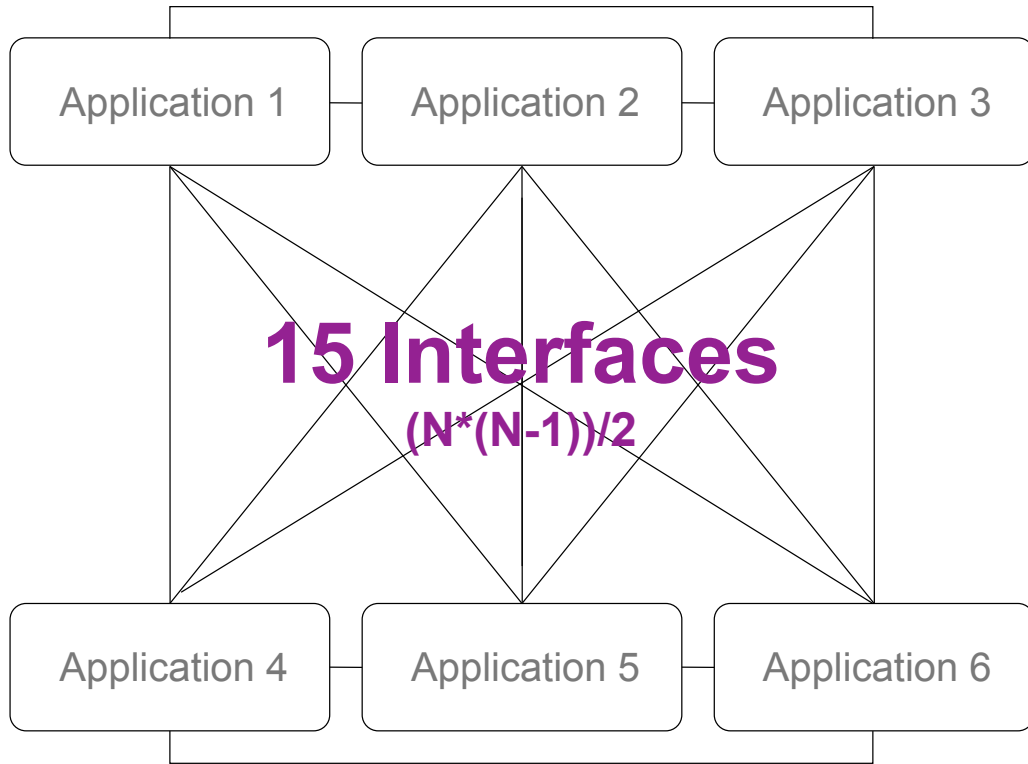


An organization's data architecture ...



... maps between and across software packages

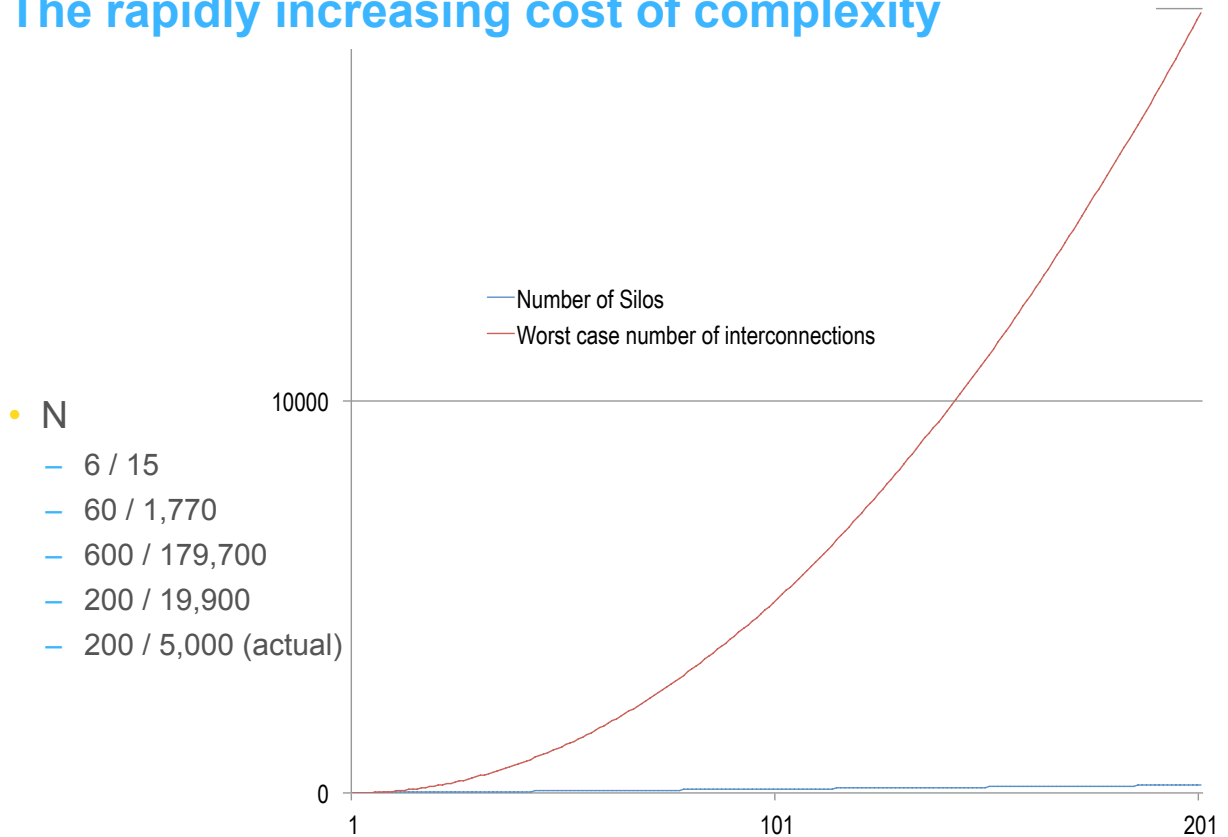
How many interfaces are required to solve this integration problem?



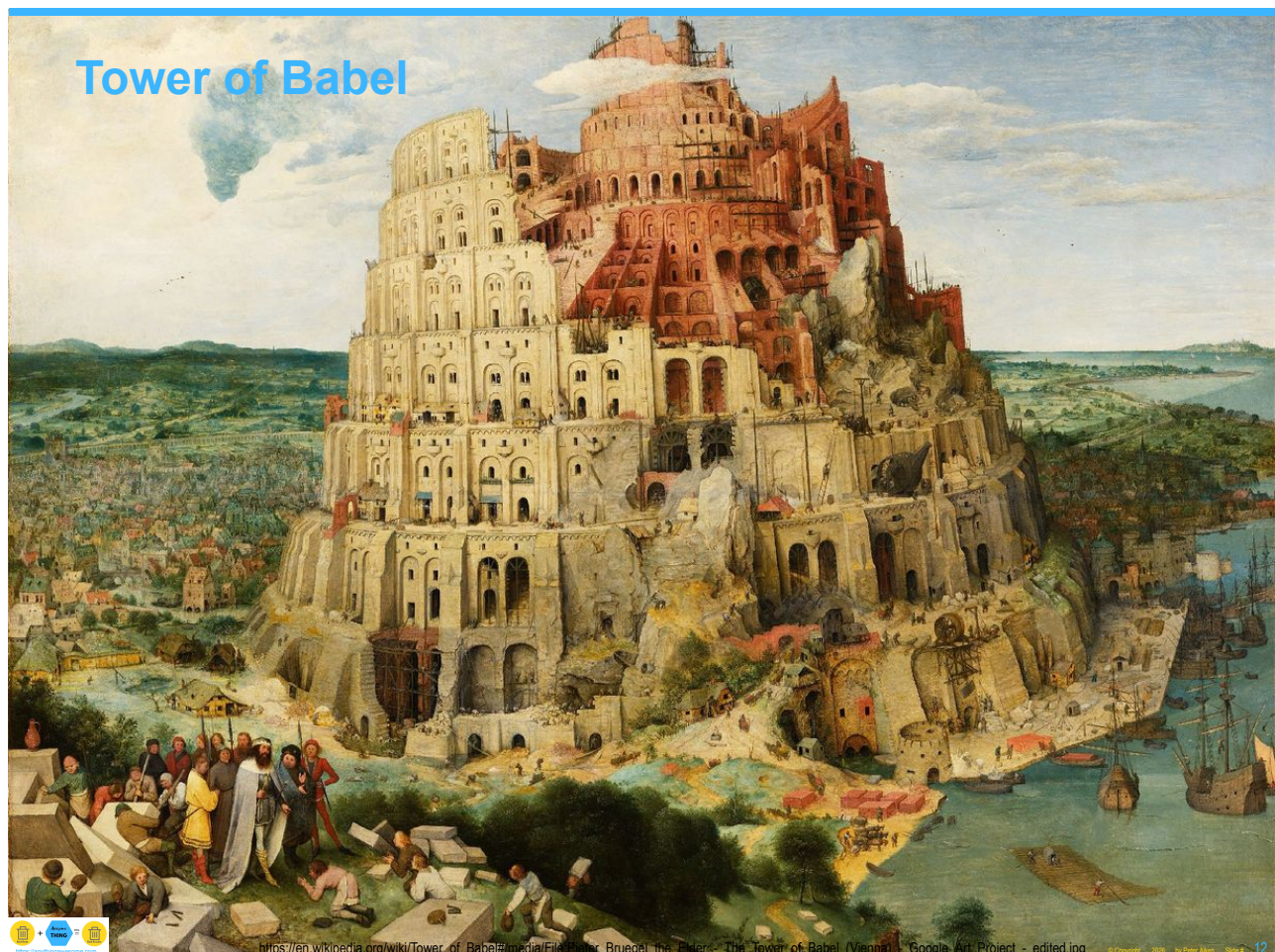
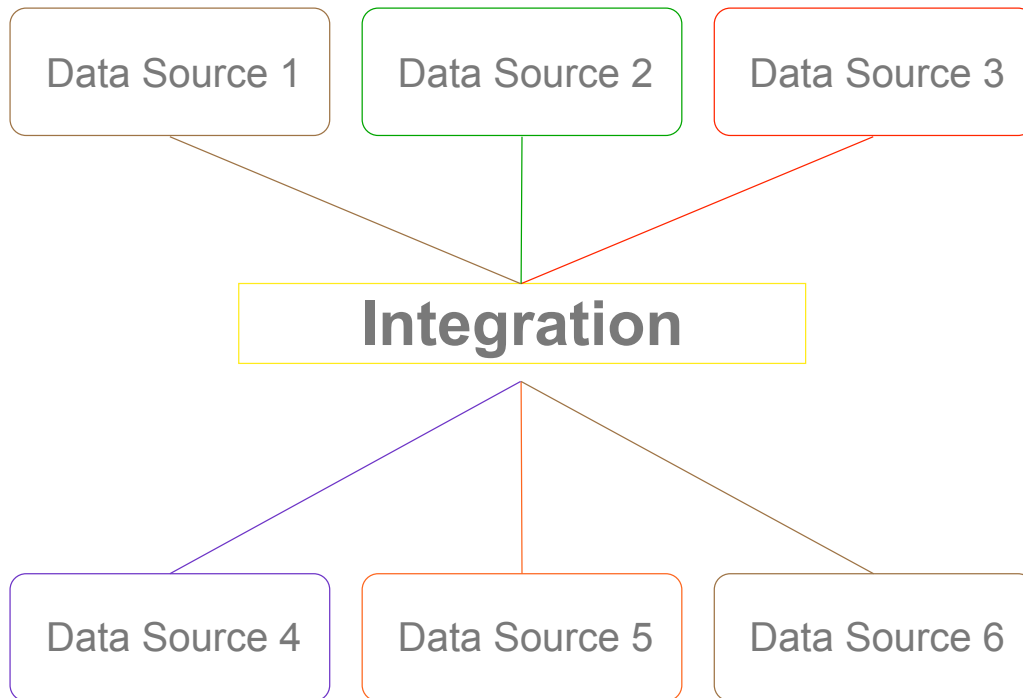
RBC: 200 applications - 4900 batch interfaces



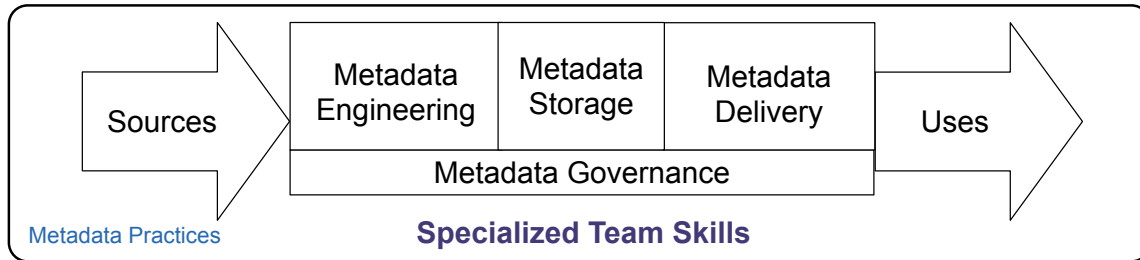
The rapidly increasing cost of complexity



Powerpoint-based Integration Solution



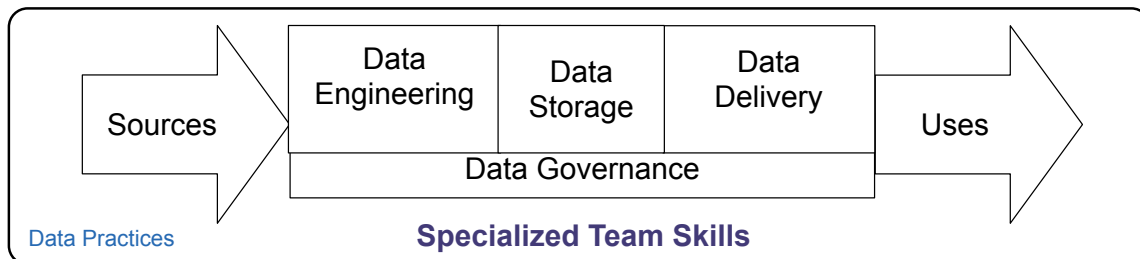
What is the structure of metadata practices?



- Metadata practices connect data sources and uses in an organized and efficient manner
 - Storage: repository, glossary, models, lineage - currently multiple technologies are used
 - Engineering: identifying/harvesting/normalizing/administer evolving metadata structures
 - Delivery: supply/access/portal/definition/lookup search identify/ensure required metadata supplies to meet business needs
 - Governance: ensure proper/creation/storage/integration/control to support effective use
- When executed, engineering and delivery implement governance



What is the structure of data practices?

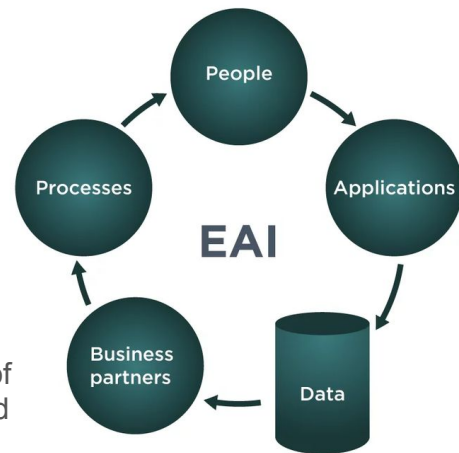
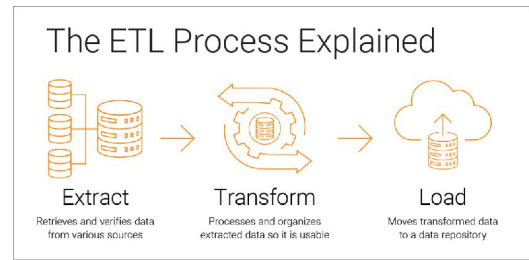


- Data practices connect data sources and uses in an organized and efficient manner
 - Storage: repository, glossary, models, lineage - currently multiple technologies are used
 - Engineering: identifying/harvesting/normalizing/administer evolving metadata structures
 - Delivery: supply/access/portal/definition/lookup search identify/ensure required metadata supplies to meet business needs
 - Governance: ensure proper/creation/storage/integration/control to support effective use
- When executed, engineering and delivery implement governance



Defining Spaces

- ETL Extract Transform, Load
 - delivers aggregated data to a new database
- EAI Enterprise Application Integration
 - connects applications to other applications in a predictable manner using pre-established connections
- EII Enterprise Information Integration
 - between ETL and EAI - delivers tailored views of information to users at the time that it is required



12/13/10

© Copyright 2008 by Peter Allen Slide 17

MetaBase Modeler - Transactions.xmi - MetaMatrix MetaBase Modeler

File Edit Navigate Search Project Metadata Run Window Help

MetaMatrix Integration Example

*Basic.xmi Transactions.xmi X

- **EII** Enterprise Information Integration
 - between ETL and EAI
 - delivers tailored views of information to users at the time that it is required

Diagram Editor Table Editor

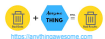
ansactions/Equities/Cusip : string

© Copyright 2008 Writable 18

Section

Summary

- Integration goals can only be achieved by ensuring the lowest level of granularity matches perfectly
- If we had it all to do over again - we wouldn't build what we currently have
- Legacy - anything that is in production
- It is reasonable to ask how qualified some decision makers are
- Integration can be easier in a software packages environment
- Complexity can be reduced but the result is always easier but rarely easy
- Implementation of appropriate standardization can be useful



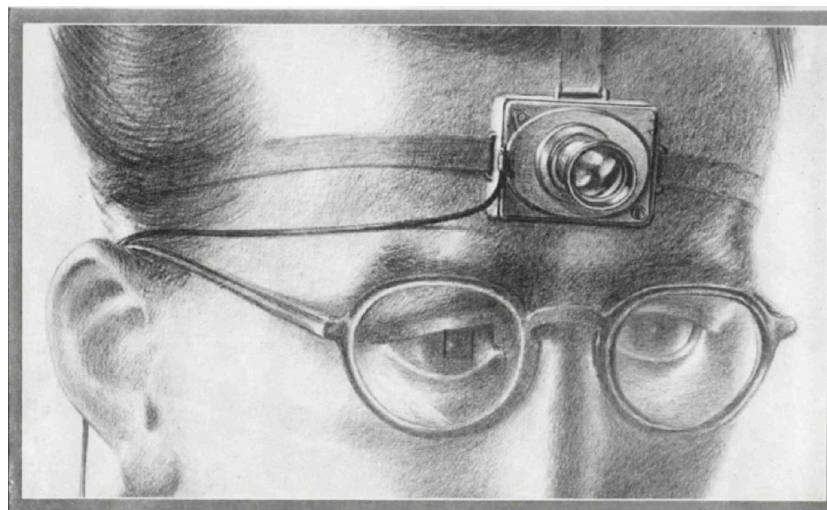
Program overview

- Integration Goals/Challenges
- AI-based approach-augmentation
- Metadata
 - Replace the concept of technical detail with that of a foundational language (common vocabulary) that provides context and clarity for your data and AI
- Cloud
 - Before taking full advantage of the scalability, agility, and cost-effectiveness of cloud services, cloud data should be architected
- Data Warehousing
 - Needs a new layer to prepare for AI
- Take aways, references and Q&A

*Data Integration Demystified:
Enabling AI-Driven Interoperability*



Amplify / Augment / Elevate / Enrich / Enhance / Fortify / Refine



A SCIENTIST OF THE FUTURE RECORDS EXPERIMENTS WITH A TINY CAMERA FITTED WITH UNIVERSAL-FOCUS LENS. THE SMALL SQUARE IN THE EYEGLASS AT THE LEFT SIGHTS THE OBSERVATION.

AS WE MAY THINK

A TOP U. S. SCIENTIST FORESEES A POSSIBLE FUTURE WORLD IN WHICH MAN-MADE MACHINES WILL START TO THINK

by VANNEVAR BUSH

DIRECTOR OF THE OFFICE OF SCIENTIFIC RESEARCH AND DEVELOPMENT
Condensed from the *Atlantic Monthly*, July 1945

This has not been a scientists' war; it has been a war in which all have had a part. The scientists, burying their old professional competition in the demand of a common cause, have shared greatly and learned much. It has been exhilarating to work in effective partnership. What are the scientists to do next?

For the biologists, and particularly for the medical scientists, there can be little indecision, for their war work has hardly required them to leave the old paths. Many indeed have been able to carry on their war research in their familiar peacetime laboratories. Their objectives remain much the same.

It is the physicists who have been thrown most violently off stride, who have left academic pursuits for the making of strange destructive gadgets, who have devised new methods for their unanticipated assignments, their part on the devices that made it possible to turn back have worked in combined effort with the physicists of our

ress, and the effort to bridge between disciplines is correspondingly superficial.

Professionally our methods of transmitting and reviewing the results of research are generations old and by now are totally inadequate for their purpose. If the aggregate time spent in writing scholarly works and in reading them could be evaluated, the ratio between these amounts of time might well be startling. Those who conscientiously attempt to keep abreast of current thought, even in restricted fields, by close and continuous reading might well shy away from an examination calculated to show how much of the previous month's efforts could be produced on call.

Mendel's concept of the laws of genetics was lost to the world for a generation because his publication did not reach the few who were capable of grasping and extending it. This sort of catastrophe is undoubtedly being repeated all about us as truly significant attainments become lost in the mass of the inconsequential.

Vannevar Bush's article, found on pages 101-108 of *The Atlantic*, is important in the history of technology as it anticipates several major aspects of personal computing as we know it today, including hypertext, the internet, and online encyclopedias. His proposed 'Memex' was a system for storing, organizing, and retrieving vast amounts of data, consisting of a desk equipped with projection screens, keyboards, and levers that offered access to microfilmed texts and imagery. Though today accomplished in the digital realm with computers and software, the general concepts put forth by Bush have revolutionized everyday life through the widespread use of services like Google and Wikipedia.

[Source: <https://www.rauction.com/auctions/lot-detail/348350206903145-vannevar-bush-as-we-may-think-in-the-atlantic-july-1945/#mz-expanded-view-141374581666j>]





Gemini



☰

🔍

📄 New chat

📁 My Stuff

📁 Gems

📄 Data Governance Professional

Chats

Fix Outlook Mac Google Sync Issues

please create an infographic of the fo...

Digital Business Architecture Quiz Cr...

Convert Canvas Quizzes to New Quiz...

I will be working with the following hi...

Safely Destroying Hard Drive Data

Please make a band logo using this i...

Virginia Data Center Market Overview

📄 ...nding Strategy: Definition, L...

Gemini



My Stuff

Documents

- 📄 **Chapter: AI and Data Debt**
Chapter Z: AI and Data Debt – The Unseen Impediment to Intelligent Systems The promise of Artificial Intelligence (AI) and Machine Learning (ML) is nothing short of revolutionary. From...
- 📄 **Chapter: Analytics and Data Debt**
Chapter Y: Analytics and Data Debt – The Unseen Cost of Insight The pursuit of data-driven insights has become a cornerstone of modern business strategy. Organizations invest heavily in...

Media



☰
🔍
Gemini
🌟

📄 New chat
🗑️

My Stuff
>

Gems
>

🔗 Data Governance Professional

Chats

Fix Outlook Mac Google Sync Issues

please create an infographic of the fo...

Digital Business Architecture Quiz Cr...

Convert Canvas Quizzes to New Quiz...

I will be working with the following hi...

Safely Destroying Hard Drive Data

Please make a band logo using this i...

Virginia Data Center Market Overview

🔗
ending Strategy: Definition, L...

My Stuff

Documents

📄 **Chapter: AI and Data Debt**

Chapter Z: AI and Data Debt – The Unseen Impediment to Intelligent Systems The promise of Artificial Intelligence (AI) and Machine Learning (ML) is nothing short of revolutionary. From...

📄 **Chapter: Analytics and Data Debt**

Chapter Y: Analytics and Data Debt – The Unseen Cost of Insight The pursuit of data-driven insights has become a cornerstone of modern business strategy. Organizations invest heavily in...

Media

☰
🔍
Gemini
🌟

📄 New chat
🗑️

My Stuff
>

Gems

🔗 Data Governance Professional

Chats

Fix Outlook Mac Google Sync Issues

please create an infographic of the fo...

Digital Business Architecture Quiz Cr...

Convert Canvas Quizzes to New Quiz...

I will be working with the following hi...

Safely Destroying Hard Drive Data

Please make a band logo using this i...

Virginia Data Center Market Overview

🔗
ending Strategy: Definition, L...

My Stuff

Documents

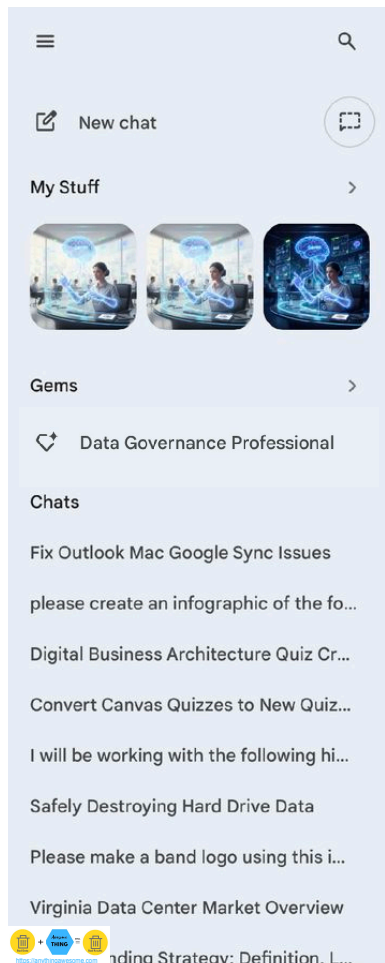
📄 **Chapter: AI and Data Debt**

Chapter Z: AI and Data Debt – The Unseen Impediment to Intelligent Systems The promise of Artificial Intelligence (AI) and Machine Learning (ML) is nothing short of revolutionary. From...

📄 **Chapter: Analytics and Data Debt**

Chapter Y: Analytics and Data Debt – The Unseen Cost of Insight The pursuit of data-driven insights has become a cornerstone of modern business strategy. Organizations invest heavily in...

Media



please explain gem

In the context of the Google ecosystem and modern technology, **Gem** usually refers to one of two things: a customizable version of the Gemini AI, or a specific shorthand in data management.

1. Gemini "Gems" (The AI Tool)

Gems are custom versions of Gemini that you can create to act as specialists for specific tasks. Instead of giving the AI a long set of instructions every time you start a new chat, you can "program" a Gem with a specific personality, knowledge set, and set of goals. [🔗](#) [🔗](#)

- **Customization:** You can give a Gem a name and a specific set of instructions (e.g., "Act as a technical editor for data management manuscripts"). [🔗](#)
- **Consistency:** Every time you use that Gem, it remembers the tone and rules you established, ensuring a consistent output. [🔗](#)
- **Use Cases:** People often create Gems for coding assistance, brainstorming creative writing, or acting as a specialized tutor for a specific subject.

2. "Gem" in Data Management

In some technical and data architecture circles, "GEM" can stand for **Generalized Enterprise Model**. This refers to a high-level conceptual framework used to describe the fundamental structures and processes of an organization.

It is often used as a starting point for data modeling to ensure that the data architecture aligns with the actual business capabilities and goals of the enterprise.

27

Prompt engineering with GenAI



write me a prompt to get you to respond to my questions as an experienced data governance professional

◆ Here is a prompt you can use:

"From now on, respond to my questions as an **experienced Data Governance professional**. Your answers should be **authoritative, practical, and strategic**, reflecting a deep understanding of data quality, compliance, policy development, metadata management, and the organizational impact of a robust data governance framework. Use terminology appropriate for a professional setting. Do not break character."



VCU Policy Program Home

Home > Active University Policies

Academic Affairs

Administration and Operations

Athletics

Board of Visitors Policies

Development and Alumni Relations

Faculty Affairs

Finance

Human Resources

Information Technology

Research

Student Life

Welcome to VCU's Policy Program

This page houses both the policy library which contains the most current version of VCU policies and resources for developing and revising policies at VCU. This resource is provided as a service to members of the VCU community and the general public. If you encounter any issues accessing pages on this site please contact the Policy Program at policy@vcu.edu or by calling 804-828-2336.

Please note: some policies may reference other VCU policies. Policies referenced are accessible by clicking the green buttons at the bottom of the screen when any given policy is open.

Please use the search box below to search policies by keyword.



Search for documents

Found 162 record(s)

Sources

- + Add source
- Select all sources
- Academic Engagement With...
- Academic Rights & Responsi...
- Accessibility and Reasonabl...
- Additional Sources of Acade...
- Alcohol and Other Drugs.pdf
- Animals on University Prope...
- Annual Assessment of Facul...
- Assessing Student Learning ...
- Assessing Student Learning ...
- Awarding Honorary Degrees...
- Campus Expression and Spa...
- Changing the Designation o...
- Commencement Participatio...
- Commencement Participatio...
- Commencement Participatio...

Chat

VCU Policy

50 sources

These documents detail Virginia Commonwealth University's policies and procedures, covering diverse aspects of university operations. They address student academic engagement and conduct, faculty roles and responsibilities, accessibility and accommodations for disabilities, credit transfer and honorary awards, campus expression and space usage, computer and network resource use, corporate research agreements, policy creation and maintenance, criminal conviction investigations, data and information governance, debt management, deceased student refunds, reporting misconduct and retaliation protection, consensual employee-student relationships, financial aid adjustments, fraud identification and reporting, hazing prevention and discipline, the university's honor system, international travel, the learning management system, graduate handbook guidelines, military-affiliated student policies, nondiscrimination and affirmative action, parental notification for dependent students' mental health treatment, PhD program guidelines, and textbook adoption and sales, as well as visa sponsorship. Many policies emphasize compliance with federal and state laws. The documents provide definitions, specifics, procedures, and frequently asked questions to clarify expectations and ensure adherence to regulations.

- Save to note
- Add note
- Audio Overview
- Briefing doc

Studio

- Audio Overview
- Video Overview
- Mind Map
- Reports
- Flashcards
- Quiz
- Infographic
- Slide Deck
- Data Table

- + Add note
- Study guide
- Briefing doc
- FAQ
- Timeline
- VCU Policy Summary
Okay, here's a detailed briefing document summarizing the main themes and important...
- VCU Policy Timeline and Stakeholders
Okay, here is a detailed timeline and cast of characters based on the provided sources...
- VCU Student FAQs
FAQs How can students earn academic credit for prior learning or experiences outside of...
- VCU Policy Summary
Okay, here's a detailed briefing document summarizing the main themes and important...
- New note
Academic Policies and Resources at VCU: FAQ
1 What are the different ways I can earn



Section

Summary

- Pay for the \$20/month version of your favorite
- Use it regularly
- Try out new things
- Exchange experiences with others
- Identify useful sources of additional information



© Copyright 2020 by Peter Allen Slide # 33

Program Overview

- Integration Goals/Challenges
- AI-based approach-augmentation
- Metadata
 - Replace the concept of technical detail with that of a foundational language (common vocabulary) that provides context and clarity for your data and AI
- Cloud
 - Before taking full advantage of the scalability, agility, and cost-effectiveness of cloud services, cloud data should be architected
- Data Warehousing
 - Needs a new layer to prepare for AI
- Take aways, references and Q&A

*Data Integration Demystified:
Enabling AI-Driven Interoperability*

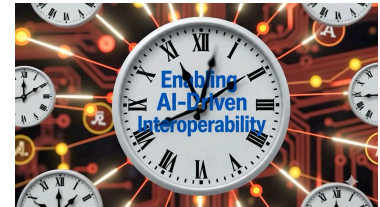
 **DATA-ED**
ONLINE
 **DATAUNIVERSITY**



Program Overview

*Data Integration Demystified:
Enabling AI-Driven Interoperability*

- Integration Goals/Challenges
- AI-based approach-augmentation
- Metadata
 - Replace the concept of technical detail with that of a foundational language (common vocabulary) that provides context and clarity for your data and AI
- Cloud
 - Before taking full advantage of the scalability, agility, and cost-effectiveness of cloud services, cloud data should be architected
- Data Warehousing
 - Needs a new layer to prepare for AI
- Take aways, references and Q&A



Data Management Body of Knowledge (DAMA DM BoK® V2)



Practice Areas

Metadata Management



The prefix meta-

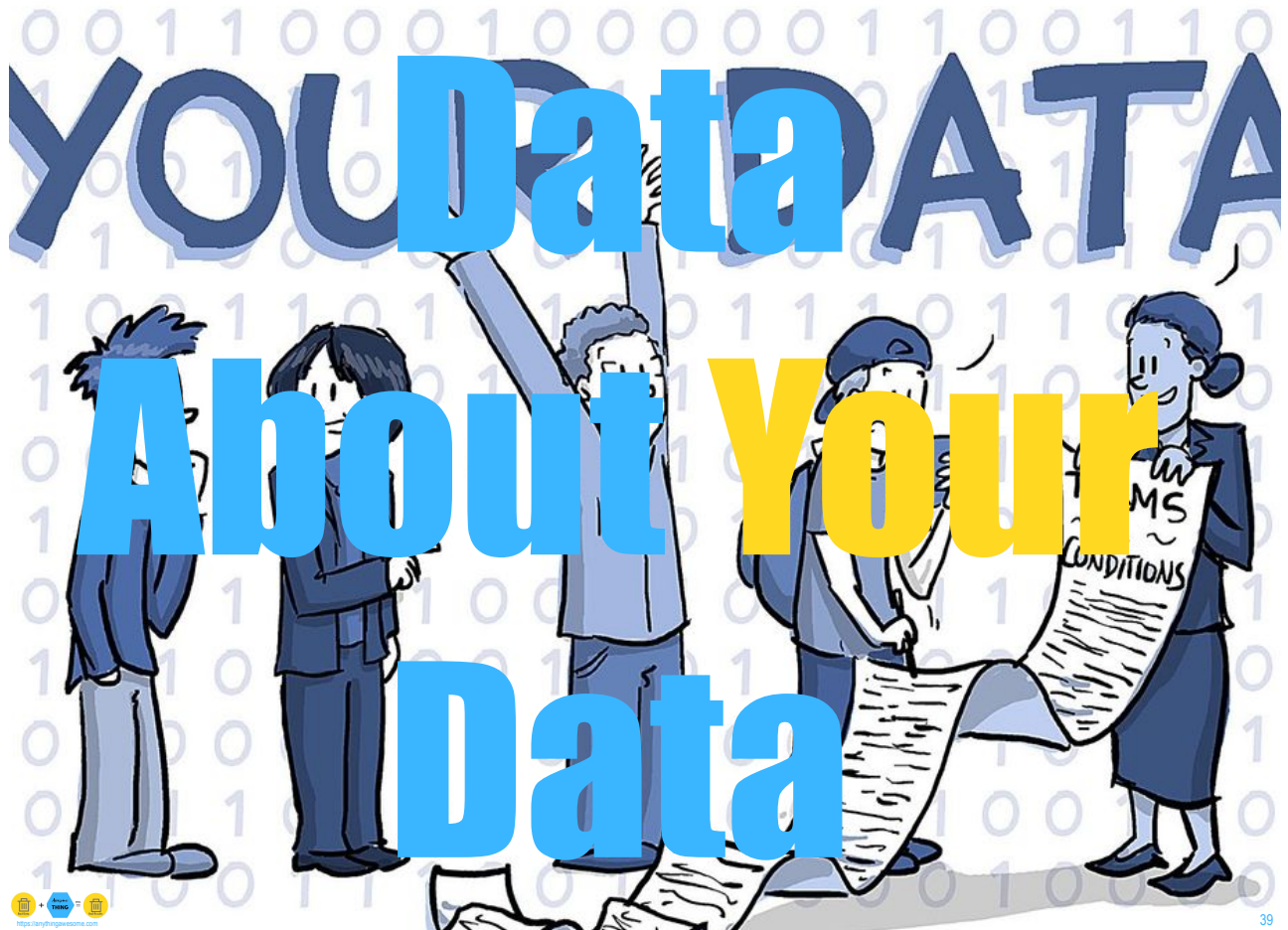
1. Situated behind: metacarpus.
2.
 - a. Later in time: metestrus.
 - b. At a later stage of development: metanephros.
3.
 - a. Change; transformation: metachromatism.
 - b. Alternation: metagenesis.
4.
 - a. Beyond; transcending; more comprehensive: metalinguistics.
 - b. At a higher state of development: metazoan.

Meta

Definition of the prefix *meta-* (Emphasis added – source: *American Heritage English Dictionary* © 1993 Houghton Mifflin).

Metadata unlocks the value of data, and therefore requires management attention [Gartner 2011]!





Using metadata descriptions of Bluetooth devices

Bluetooth

Show All

"mine's i7" is the name Bluetooth devices use to connect to this computer.

On Discoverable

CGL Trackpad
Connected

IDR Trackpad
Not Connected

Motorola S9
Not Connected

Peter's i4
Not Connected

VCU Keyboard
Not Connected

VCU Trackpad
Not Connected

Data Column	Attributes/Fields	
CGL Trackpad	Keyboard	VCU
CGL Motorola S10	Motorola S10	CGL
IDR Trackpad	Peter's i5s	
Peter's i5s	Trackpad	CGL
VCU Keyboard	Trackpad	IDR
VCU Trackpad	Trackpad	VCU

Show Bluetooth status in the menu bar

Sharing Setup... Advanced... ?

Example:
iTunes → Music™/.mp3 player app

Name	Time	Artist	Album	Genre
Track 01	2:34			
Track 02	3:14			
Track 03	3:20			
Track 04	3:12			
Track 05	2:36			
Track 06	2:49			
Track 07	3:10			
Track 08	3:02			
Track 09	3:07			
Track 10	2:18			
Track 11				
Track 12				
Track 13				
Track 14				
Track 15				
Track 16				
Track 17				
Track 18				
Track 19				
Track 20				
Track 21				
Track 22	4:24			
Track 23	4:49			
Track 24	3:47			
Track 25	4:23			

25 songs, 1.3 hours, 801.7 MB

- Example:
 - Music Metadata
- Insert a recently purchased CD
- Music can:
 - Count the number of tracks (25)
 - Determine the length of each track

Example:
Music Metadata

Name	Time	Artist	Album	Genre
Move	2:34	Miles Davis	The Complete Birth Of The Cool	Jazz
Jeru	3:14	Miles Davis	The Complete Birth Of The Cool	Jazz
Moon Dreams	3:20	Miles Davis	The Complete Birth Of The Cool	Jazz
Venus De Milo	3:12	Miles Davis	The Complete Birth Of The Cool	Jazz
Budo	3:26	Miles Davis	The Complete Birth Of The Cool	Jazz
Deception				
Godchild				
Boplicity				
Rocker				
Israel				
Rouge				
Darn That Dream				
Birth Of The Cool...				
Symphony Sid Ann...				
Move (Live)				
Why Do I Love You...				
Godchild (Live)				
Symphony Sid Intr...				
S'il Vous Plait [Live]				
Moon Dreams [Live]				
Budo (Hallucinatio...				
Darn That Dream [...]				
Move (Live)				
Moon Dreams (Live)				
Budo (Hallucinatio...				

25 songs, 1.3 hours, 801.7 MB

- When connected to the Internet the Music app connects to the Gracenote(.com) Media Database and retrieves:
 - CD Name
 - Artist
 - Track Names
 - Genre
 - Artwork
- Sure would be a pain to type in all this information (How many of your knowledge workers are suffering?)

Example: Music Metadata

- To organize my Music library
 - I create a "New Smart Playlist" for Artist's containing "Miles Davis"

Name	Time	Artist	Album	Genre
1 ✓ Move	2:34	Miles Davis	The Complete Birth Of The Cool	Jazz
2 ✓ Jeru	3:14	Miles Davis	The Complete Birth Of The Cool	Jazz
20 ✓ Moon Dreams [Live]	3:08	Miles Davis	The Complete Birth Of The Cool	Jazz
21 ✓ Budo (Hallucinatio...	3:26	Miles Davis	The Complete Birth Of The Cool	Jazz
22 ✓ Darn That Dream [...]	4:24	Miles Davis	The Complete Birth Of The Cool	Jazz
23 ✓ Move (Live)	4:49	Miles Davis	The Complete Birth Of The Cool	Jazz
24 ✓ Moon Dreams (Live)	3:47	Miles Davis	The Complete Birth Of The Cool	Jazz
25 ✓ Budo (Hallucinatio...	4:23	Miles Davis	The Complete Birth Of The Cool	Jazz

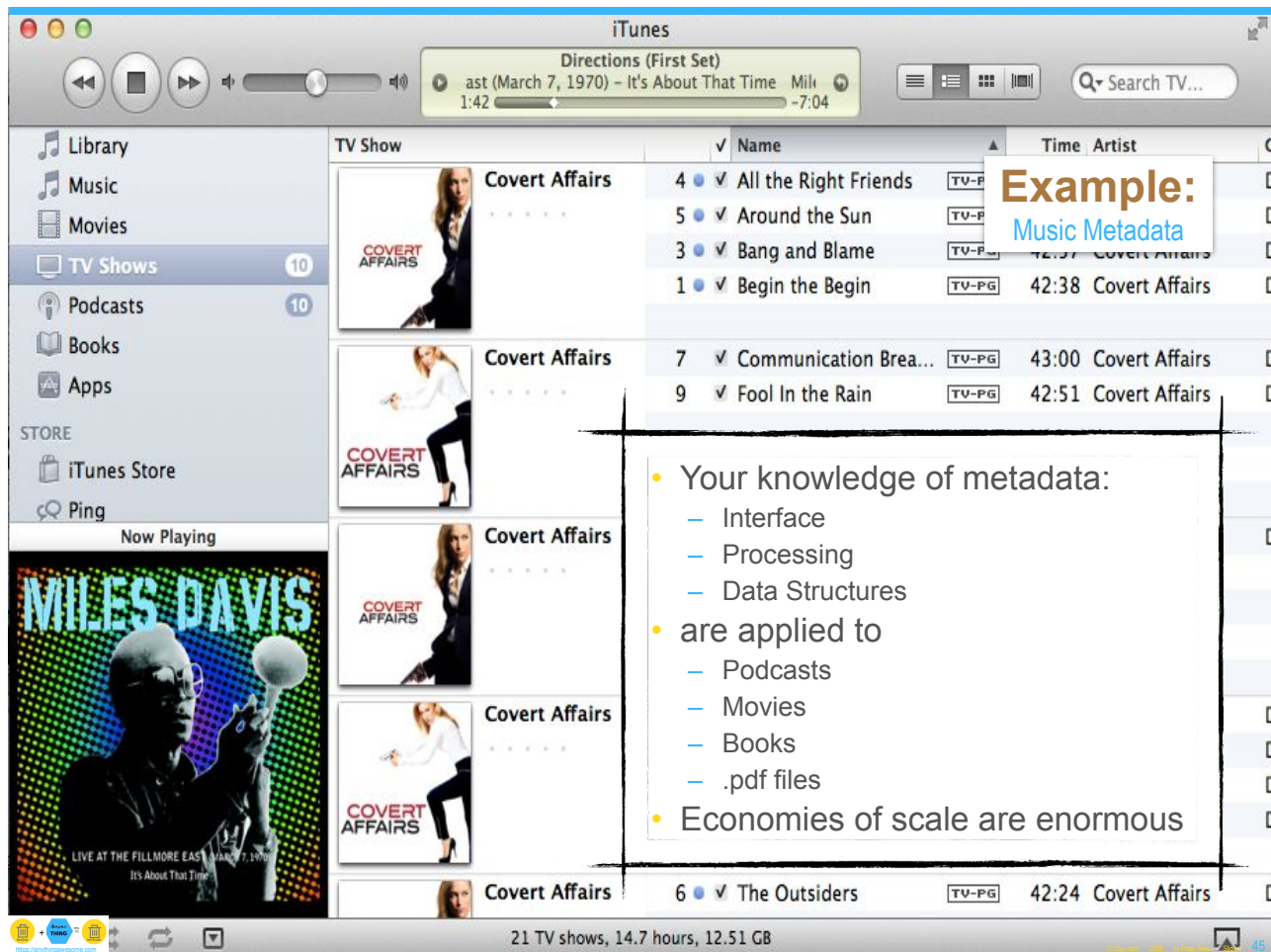
25 songs, 1.3 hours, 801.7 MB

Example: Music Metadata

- Notice I didn't get the desired results
- I already had another Miles Davis recording, "Live at the Fillmore East"
- Must fine-tune the smart playlist request to get the desired results
 - Now specify that Album must also contain "The complete birth of the cool"
- Now I can move the smart playlist "Miles Davis" to a folder
- Or not?

Date Added	Bit Rate	Time	Artist	Album
9/5/11 6:48 PM	128 kbps	3:25	Miles Davis	The Co
9/5/11 6:48 PM	128 kbps	3:15	Miles Davis	The Co
9/5/11 6:48 PM	128 kbps	2:18	Miles Davis	The Co
9/5/11 6:48 PM	128 kbps	3:07	Miles Davis	The Co

32 songs, 2.3 hours, 133.2 MB



input to this was previous 5 slides & “illustrate the concept of 'metadata' as a use of data and not as a type of data”

1. Data in Isolation (Physical vs. Logical Presence)

When you insert a physical audio CD into a computer without an internet connection, the system can only read the raw structural properties physically printed or encoded on the disc. The software can instantly extract specific technical data points:

- Track Count:** For example, determining there are exactly 25 tracks on the disc.
- Track Duration:** Measuring the precise playtime of each individual item (e.g., 2:34 or 3:14).
- Total Volume:** Calculating aggregate metrics like total run time (1.3 hours) and file size (801.7 MB).

At this stage, the actual identity of the content is completely missing. Without context, the system simply applies sequential, generic placeholders labeled “Track 01”, “Track 02”, etc. The structural data is present, but it isn't yet being *used* to describe anything meaningful to a human observer.

2. Metadata as a “Use” (Contextual Assignment)

The transformation happens when the system connects to an external database (such as Gracenote). The application sends the disc's hardware signatures (the track count and precise timings) as a query.

The database matches those physical metrics against its records and returns an administrative layer of information:

[Raw Physical Data]	[Gracenote Database Query]
Track 1 (2:34)	-----> Matches Record: "Miles Davis"
Track 2 (3:14)	-----> Matches Record: "Jeru"
Total: 25 Songs	-----> Matches Album: "The Complete Birth Of The Cool"

Once downloaded, these standard data strings function as **metadata** because of *how they are used*. They are applied directly to the files to describe their identity:

- Identity Attributes:** Overwriting generic track numbers with true names like *Jeru*, *Moon Dreams*, or *Venus De Milo*.

- Classification Attributions:** Assigning the Artist (**Miles Davis**), Album (**The Complete Birth Of The Cool**), and Genre (**Jazz**). The text string "Miles Davis" is fundamentally just text data. However, it serves a metadata function here because its **purpose** is to define who created the audio file.

3. Operationalizing Metadata (Smart Playlists)

Because this descriptive information is structured uniformly, you can build logical systems to automate tasks. For example, if you want to organize a library, you can create a **Smart Playlist** with a conditional rule instructing the application to find instances where the **Artist field contains "Miles Davis"**.

However, relying on a single descriptive attribute can sometimes pull in unintended matches if the criteria are too broad. In this case, a basic search for the artist might return tracks from entirely different recordings, such as a concert album like "*Live at the Fillmore East*".

To fix this, you must fine-tune the rule by compounding the metadata fields. By specifying that the **Artist contains "Miles Davis"** AND the **Album contains "The Complete Birth Of The Cool"**, the system filters out the surrounding noise to isolate the exact target tracks.

4. The Scale of Metadata Applicability

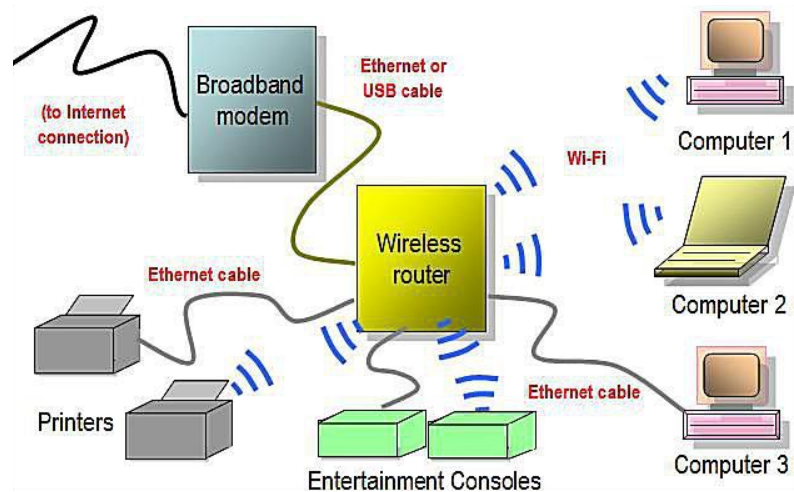
This operational principle is not exclusive to music files. The exact same foundational mechanisms—**user interfaces, backend processing, and underlying data structures**—are applied uniformly across completely different media landscapes:

- TV Shows & Movies:** Tracking titles (*Covert Affairs*), ratings (TV-PG), and episode runtimes (42:38).
- Other Media:** Standardizing organizational rules for Podcasts, Books, and standalone .pdf documents.

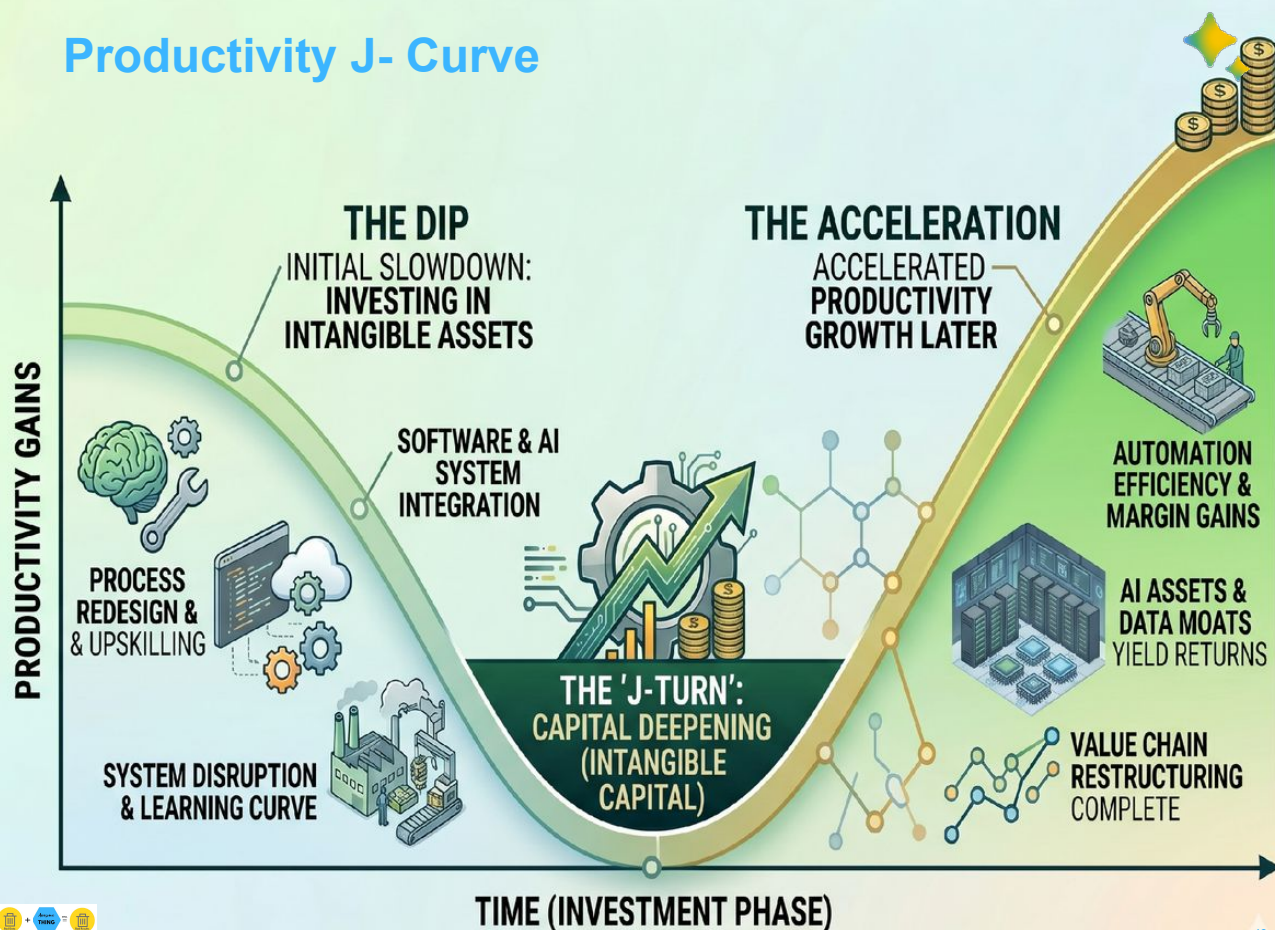
Ultimately, metadata is not a special sub-category of code or a rare breed of file; it is any data utilized to describe, locate, and control another piece of information. When applied systematically, the resulting structural economies of scale are enormous.

The most likely managed metadata in your organization

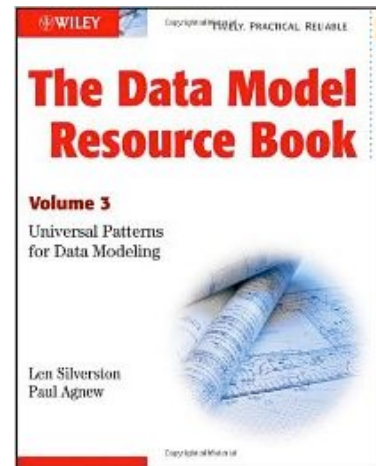
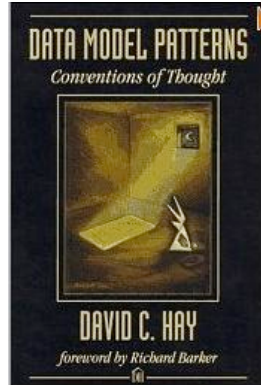
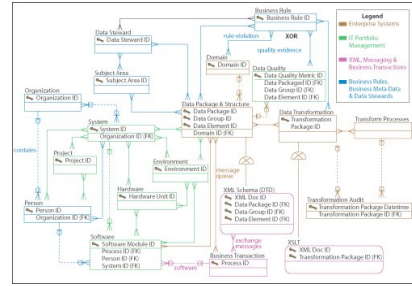
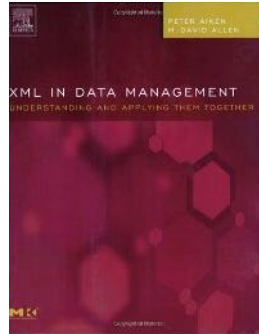
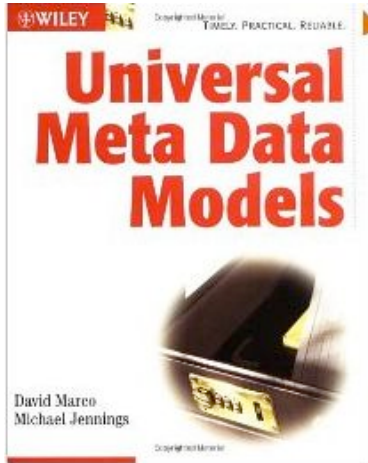
- Tracking network users and access points is metadata
- Your organization's networking group allocates the responsibility for knowing (at least):
 - All the devices permitted to logon to your network
 - Locations of all permitted access points
- This responsibility belongs to a named individual(s)



Productivity J- Curve



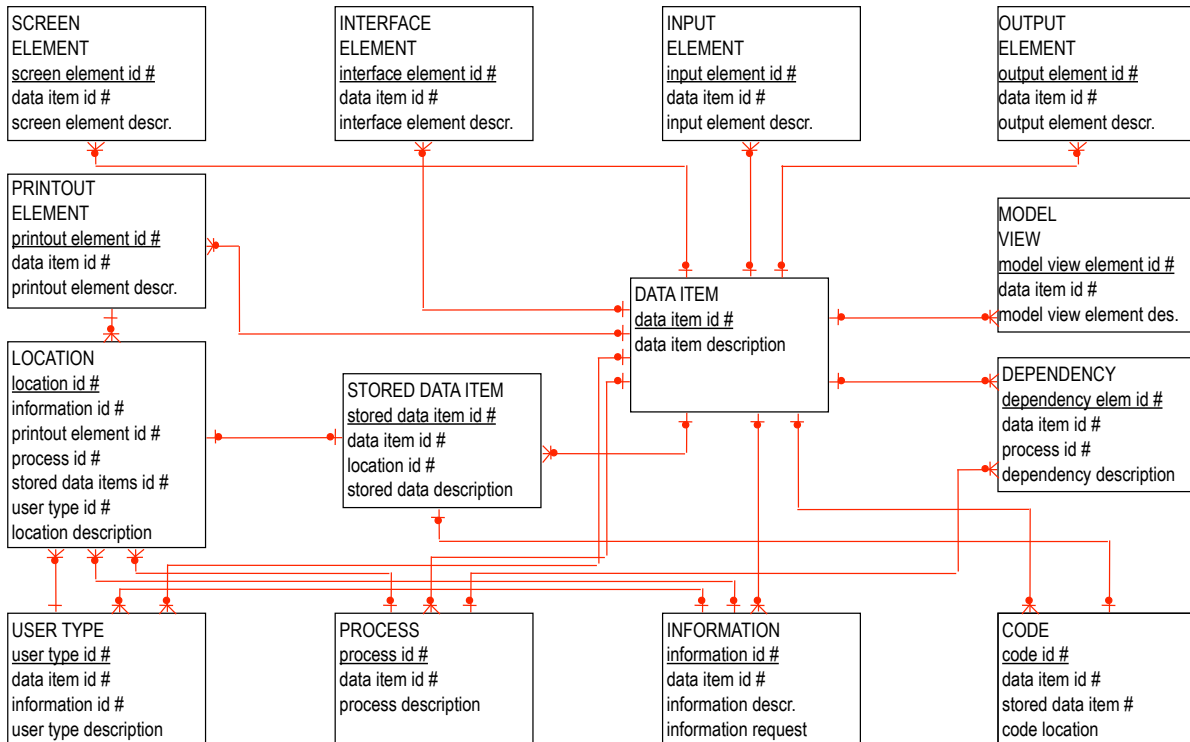
Metamodels References

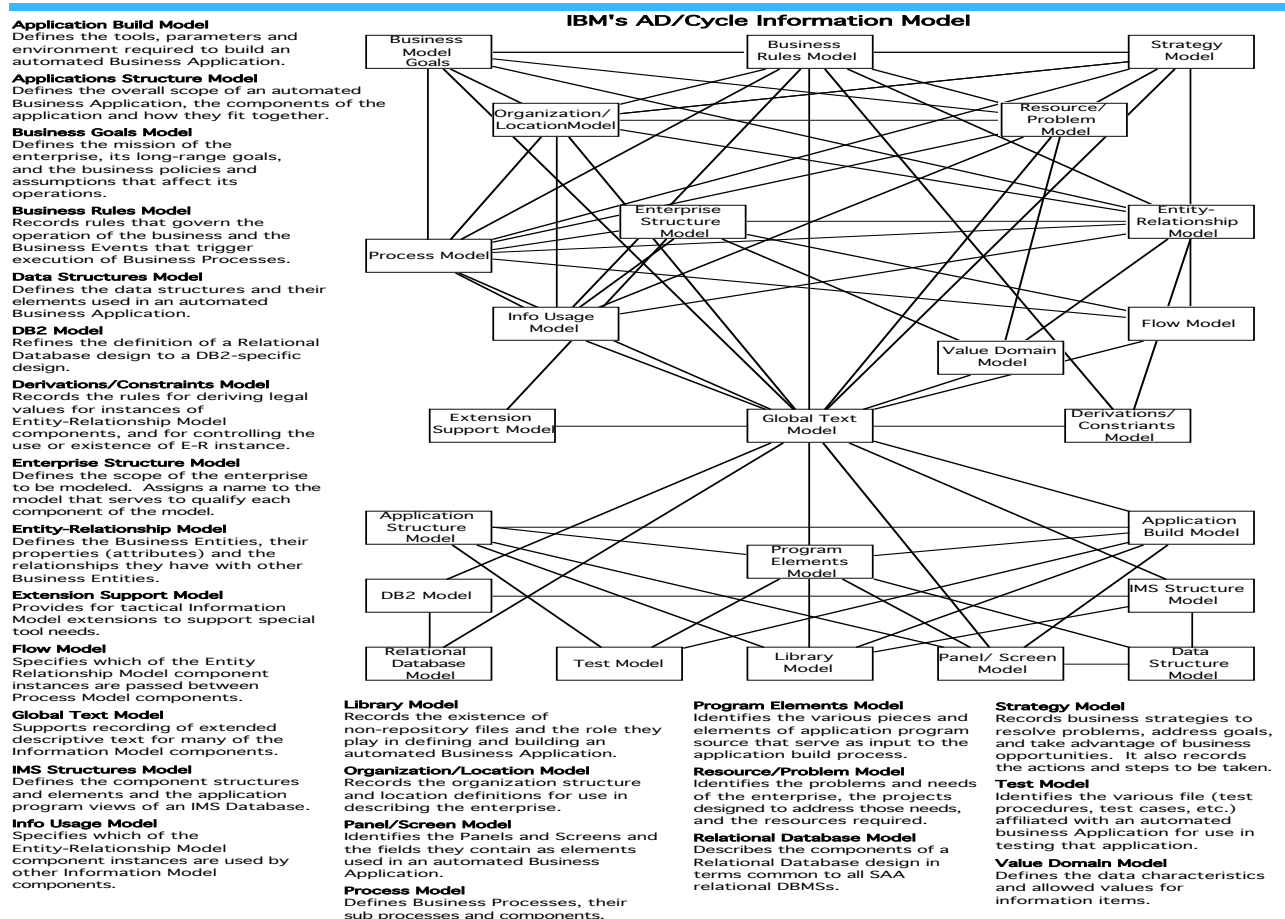


Source: http://dmreview.com/article_sub.cfm?articleID=1000941 used with permission



Metamodel





CASE-based XML Support



The screenshot displays the Visible Advantage interface for 'XYZ, Inc.' with two main windows: 'Data Dictionary: Employment' and 'Data Map: Employment'.

Data Dictionary: Employment

- Entity:** PERSON
- Category:** Principal
- Nature:** Dynamic
- Authority:** Shared Create, Read, Update, Delete
- Purpose:** Captures information on all people who are associated with the XYZ Corporation.
- Attributes:** person id #, person name, last name, first name, middle initial. Callouts point to 'DM_Attribute' and 'DM_Association'.
- Associations:** APPLICANT, EMPLOYEE, PERSON ADDRESS, PERSON JOB, PERSON ORGANIZATION, PERSON ROLE, PERSON SKILL.
- Model Views:** Human Resources, Employment, Applicant Screening, Hiring an Employee, Corporate Structure, Planning Statements.
- Model Links:** [design object]: HR Advantage.PERSON, [statement]: Critical Position Staffing, [statement]: Incomplete Applications.

Data Map: Employment

```
<!ELEMENT DM_Entity ...>
<!ATTLIST DM_Entity id ... >
<!ATTLIST DM_Entity name ... >
<!ATTLIST DM_Entity category ... >
<!ATTLIST DM_Entity nature ... >
<!ATTLIST DM_Entity authority ... >
<!ATTLIST DM_Entity purpose ... >
<!ATTLIST DM_Entity created_by ... >
<!ATTLIST DM_Entity created_date_time ... >
<!ATTLIST DM_Entity modified_by ... >
<!ATTLIST DM_Entity modified_date_time ... >
```

The diagram shows relationships between entities: SKILL, PERSON, JOB, and APPLICANT. Callouts include 'DM_Model_View', 'DM_Model_Link', and 'One or Many Occurrences per Entity'.



Metadata Generation Example

The screenshot displays the Visible Advantage interface for 'XML Generation Example' with two main windows: 'Design Dictionary: XML Order Entry' and 'Design Map: XML Order Entry'.

Design Dictionary: XML Order Entry

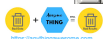
- Design Object:** CUSTOMER
- Type:** Data Structure
- Structure Type:** Not Defined
- Nature:** Dynamic
- Purpose:** Person
- Properties:** customer, customer, customer, customer
- Access Mech:** Primary
- Relationships:** CUS, ORL
- Views:** XML Orc
- Model Links:** (entity):
- Created:** Wed Ma in sessi
- Modified:** Wed Ma in session BUSINESS RE-ENGINEERING

Design Map: XML Order Entry

The diagram shows relationships between entities: NEED, CUSTOMER_NEED, PRODUCT, PRODUCT_NEED, and ORDER_PRODUCT.

Generate Schema Dialog

- Design View:** XML Order Entry
- Candidates:** Create DCD, XML Data
- Include:** Create DTD with Attlist, Attlist with Data
- Buttons:** OK, Cancel





- They know you rang a phone sex service at 2:24 am and spoke for 18 minutes. *But they don't know what you talked about.*
- They know you called the suicide prevention hotline from the Golden Gate Bridge. *But the topic of the call remains a secret.*
- They know you spoke with an health testing service, then your doctor, then your health insurance company in the same hour. *But they don't know what was discussed.*
- They know you received a call from the local NRA office while it was having a campaign against gun legislation, and then called your senators and congressional representatives immediately after. *But the content of those calls remains safe from government intrusion.*
- They know you called a gynecologist, spoke for a half hour, and then called the local Planned Parenthood's number later that day. *But nobody knows what you spoke about.*

– <https://www.eff.org/deeplinks/2013/06/why-metadata-matters>



The screenshot shows the DataEd-Integration web application. The interface is divided into three main sections: Sources, Chat, and Studio. The Sources panel on the left includes a search bar and a list of sources, with one source titled "2026 Data Integra..." selected. The Chat panel in the center displays a document snippet with the following text:










DataEd-Integration
1 source - Jun 8, 2026

The provided material explores the essential role of **data integration** in creating functional **AI-driven systems** and interoperable environments. It highlights a critical warning that even the most advanced technologies will produce **poor results** if they are fed **low-quality information**. To address these challenges, the presentation outlines how **metadata strategies**, **cloud-based technologies**, and **modern warehousing** can be leveraged to enhance system efficiency and responsiveness. The text further emphasizes the importance of establishing a clear **enterprise taxonomy** and engaging the business community to define data entities accurately. Ultimately, these efforts aim to streamline how **organizations and individuals** interact with complex digital infrastructures. This overview serves as a roadmap for moving from **fragmented data silos** toward a unified and high-performing technical architecture.










The Studio panel on the right contains various interactive elements like Audio, Video, Reports, Quizzes, and Slides, along with a list of recent documents such as "The AI Bedrock...", "AI-Driven...", "Data Integration for ...", and "The Metadata Engin..."





 Audio... >	 Slide... >
 Video... >	 Mind...
 Repor... >	 Flash... >
 Quiz >	 Infogr... >
 Data... >	



 Audio... >	 Slide... >
 Video... >	 Mind...
 Repor... >	 Flash... >
 Quiz >	 Infogr... >
 Data... >	

Audio... >	Slide... >
Video... >	Mind...
Repor... >	Flash... >
Quiz >	Infogr... >
Data... >	

生成式AI：您的数据管理新搭档





Audio... >

Slide... >

Video... >

Mind... >

Repor... >

Flash... >

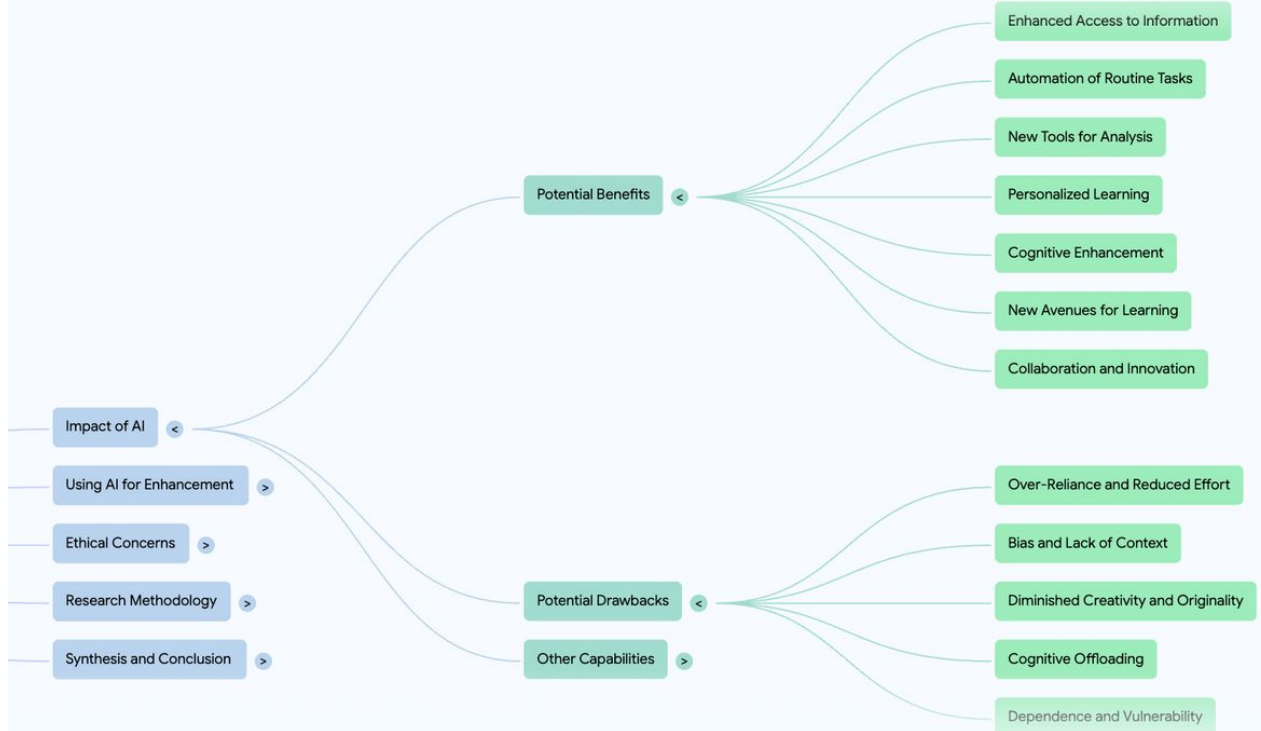
Quiz >

Infogr... >

Data... >



NotebookLM Mind Map



Audio...
>

Slide...
>

Video...
>

Mind...
>

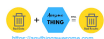
Repor...
>

Flash...
>

Quiz
>

Infogr...
>

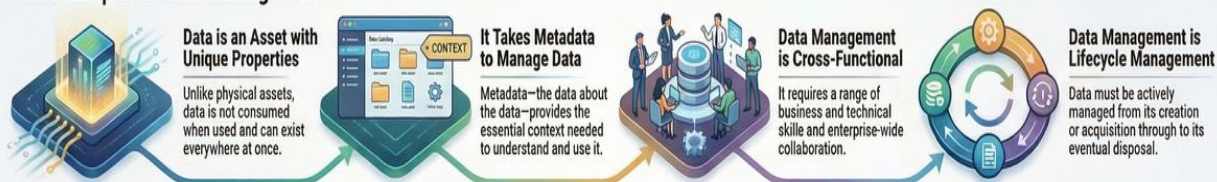
Data...
>



Managing Data as a Strategic Asset: A DAMA-DMBOK Overview

A high-level overview of core principles and framework for professional data management, treating data as an enterprise asset.

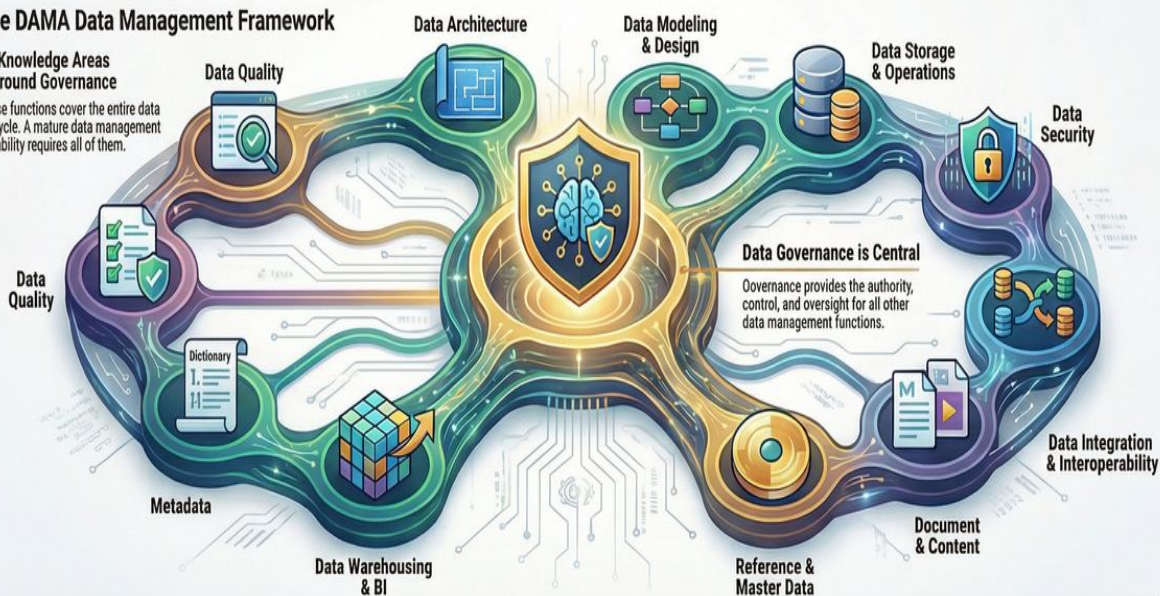
Core Principles of Data Management



The DAMA Data Management Framework

10 Knowledge Areas Surround Governance

These functions cover the entire data lifecycle. A mature data management capability requires all of them.



The Exponential Cost of Complexity: $N*(N-1)/2$



$N*(N-1)/2$

Rapidly Increasing Cost of Complexity. In a typical siloed system, the number of required interfaces explodes, for example, 200 applications can require up to 4,900 batch interfaces.

- 1 Silo: 0 Interconnections
- 6 Silos: 15 Interconnections
- 100 Silos: 4,550 Interconnections
- 200 Silos: 19,900 (Theoretical) / 4,900 (Reported Batch)

Section 1: The Common Vocabulary (Metadata as Context)

Metadata: To Data what Data is to Real Life.

Just as data reflects real-world transactions and objects, metadata reflects the data itself--providing the essential context AI needs to understand and process information.



The 'Common Vocabulary' for AI.

A 'Trusted Catalog' or business glossary provides a shared language across business, techniques, and organizational strategy.

A 'Trusted Catalog' or business glossary provides a shared language across business, technical, and system layers, ensuring data assets support organizational strategy.

Section 2: From 'Smart Playlists' to 'Smart Enterprises'



Automating Discovery at Scale. Smart Enterprises use metadata rules and 'Data Blueprints' to automate data discovery and integration, avoiding the trap of manual interconnections.



Eliminating the 'Siloed' Highway. Moving from unaligned business vocabularies and legacy debt to data centric workflows driven by a centralized Metadata Repository.

CLARITY: AI-Driven Interoperability & Efficiency



Beyond Simple Tagging. Metadata transcends basic linguistics; it is integral to every IT system, tracking everything from device access points to complex data lineages.

Section 3: Using GenAI to Scale Metadata Engineering

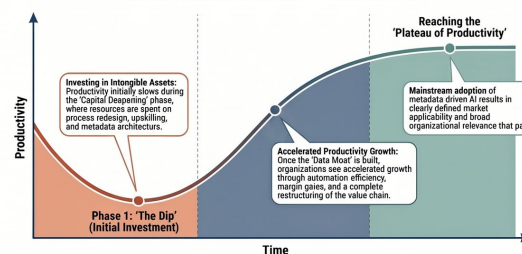


Generative AI (like Gemin) acts as a force multiplier for data professionals, aids identifying, harvesting, and normalizing evolving metadata structures.

Building the Business Glossary. AI helps scale the creation of data dictionaries and enterprise taxonomies, defining entities and providing agreed-upon business definitions across the domain.

Automated Metadata Management. AI facilitates the planning and control activities required to enable access to high-quality, integrated metadata from diverse, fragmented sources.

Section 4: The 'Productivity J-Curve' of Metadata Investment



The Metadata Engine: Powering AI-Driven Interoperability

The Exponential Cost of Complexity: $N*(N-1)/2$



$N*(N-1)/2$

Rapidly Increasing Cost of Complexity, In a typical siloed system, the number of required interfaces explodes, for example, 200 applications can require up to 4,900 batch interfaces.

- 1 Silo: 0 Interconnections
- 6 Silos: 15 Interconnections
- 100 Silos: 4,550 Interconnections
- 200 Silos: 19,900 (Theoretical) / 4,900 (Reported Batch)

Section 1: The Common Vocabulary (Metadata as Context)

Metadata: To Data what Data is to Real Life.

Just as data reflects real-world transactions and objects, metadata reflects the data itself--providing the essential context AI needs to understand and process information.



The 'Common Vocabulary' for AI.

A 'Trusted Catalog' or business glossary provides a shared language across business, techniques, and organizational strategy.

A 'Trusted Catalog' or business glossary provides a shared language across business, technical, and system layers, ensuring data assets support organizational strategy.

The Metadata Engine: Powering AI-Driven Interoperability



Section 2: From 'Smart Playlists' to 'Smart Enterprises'



Smart Playlists

Automating Discovery at Scale
Smart Enterprises use metadata rules and 'Data Blueprints' to automate data discovery and integration, avoiding the trap of manual interconnections.

VS



Data Blueprint

Eliminating the 'Siloed' Highway.
Moving from unaligned business vocabularies and legacy debt to data centric workflows driven by a centralized Metadata Repository.

CLARITY: AI-Driven Interoperability & Efficiency



Beyond Simple Tagging

Metadata transcends basic linguistics; it is integral to every IT system, tracking everything from device access points to complex data lineages.

The Metadata Engine: Powering AI-Driven Interoperability



Section 3: Using GenAI to Scale Metadata Engineering



AI-Driven Augmentation

Generative AI (like Gemini) acts as a force multiplier for data professionals, as identifying, harvesting, and normalizing evolving metadata structures.

Building the Business Glossary

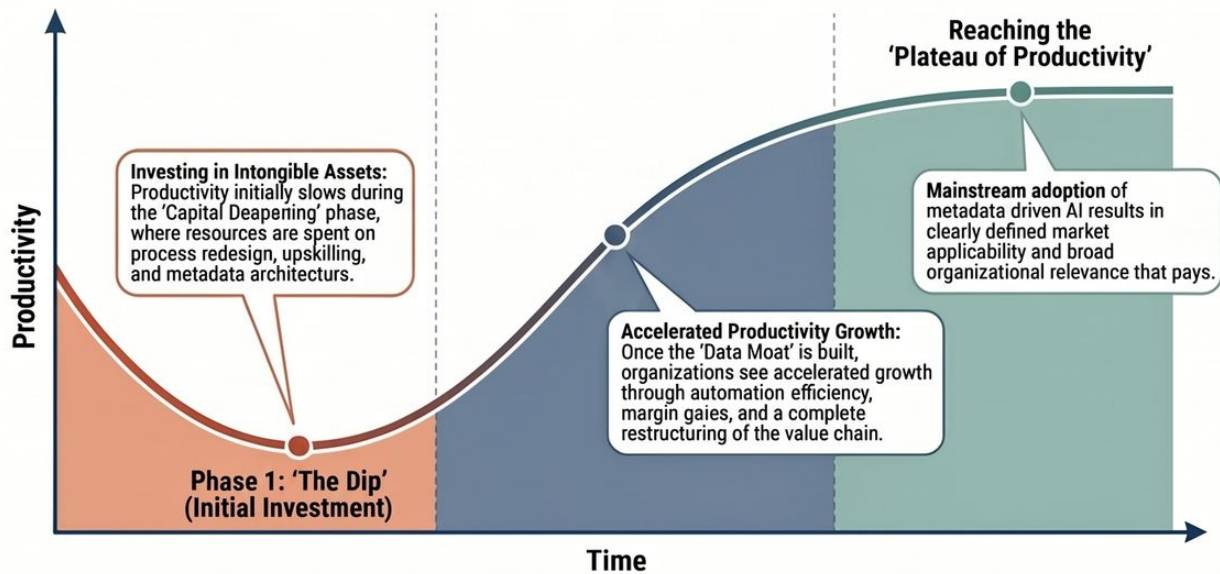
AI helps scale the creation of data dictionaries and enterprise taxonomies, defining entities and providing agreed-upon business definitions across the domain.

Automated Metadata Management

AI facilitates the planning and control activities required to enable access to high-quality, integrated metadata from diverse, fragmented sources.



Section 4: The 'Productivity J-Curve' of Metadata Investment



Program Overview

- Integration Goals/Challenges
- AI-based approach-augmentation
- Metadata
 - Replace the concept of technical detail with that of a foundational language (common vocabulary) that provides context and clarity for your data and AI

*Data Integration Demystified:
Enabling AI-Driven Interoperability*

- Cloud
 - Before taking full advantage of the scalability, agility, and cost-effectiveness of cloud services, cloud data should be architected

- Data Warehousing
 - Needs a new layer to prepare for AI
- Take aways, references and Q&A





February 26, 2010

Similar Opportunities

MEMORANDUM FOR CHIEF INFORMATION OFFICERS

FROM: Vivek Kundra
Federal Chief Information Officer

SUBJECT: Federal Data Center Consolidation Initiative

- **Adopt Light Technologies and Shared Solutions.** We are reducing our data center footprint by 40 percent by 2015 and shifting the agency default approach to IT to a **cloud-first policy** as part of the 2012 budget process. Consolidating more than 2,000 government data centers will save money, increase security and improve performance.

The reported number of Federal data centers grew from 432 in 1998 to more than 1,100 in 2009.¹ This growth in redundant infrastructure investments is costly, inefficient, unsustainable and has a significant impact on energy consumption. In 2006, Federal servers and data centers consumed over 6 billion kWh of electricity and without a fundamental shift in how we deploy technology it could exceed 12 billion kWh by 2011.² In addition to the energy impact, information collected from agencies in 2009 shows relatively low utilization rates of current infrastructure and limited reuse of data centers within or across agencies. The cost of operating a single data center is significant, from hardware and software costs to real estate and cooling costs.

- **IT Infrastructure.** Your submission should include funding for the timely execution of agency plans to consolidate data centers developed in FY 2010 (reference FY 2011 passback guidance). In coordination with the data center consolidations, agencies should evaluate the potential to adopt cloud computing solutions by analyzing computing alternatives for IT investments in FY 2012. Agencies will be expected to adopt cloud computing solutions where they represent the best value at an acceptable level of risk.



<https://builtin.com/articles/clarify-cloud-spending>

Cloud Spending Infuriates Executives. Here's How to Solve It.

While C-suite disagreements are to be expected, squabbles over the way a company spends on cloud services don't have to be.



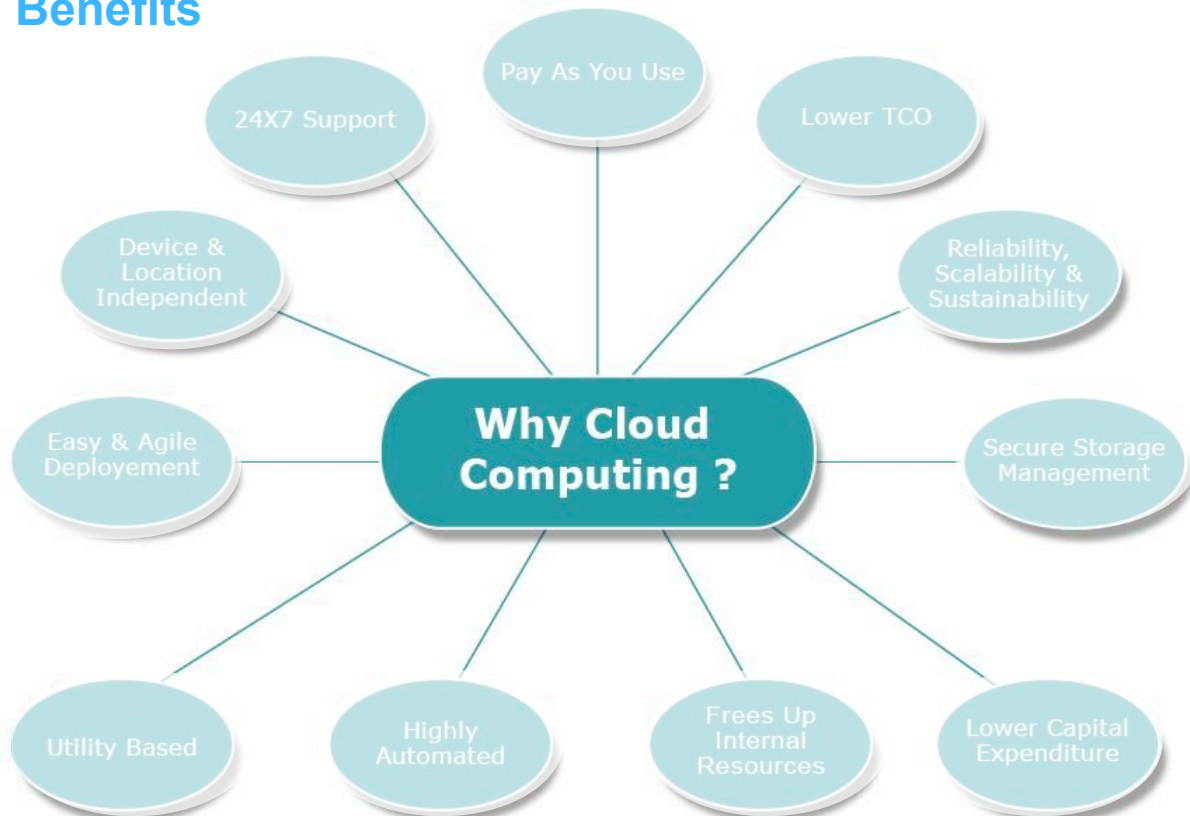
Written by [Willy Sennott](#)
Published on Aug. 24, 2023



Aspirational Data in the Cloud



Benefits



Data in the cloud should have three attributes that data outside the cloud should not have - it should be:

Cleaner



Smaller

Sharable-er

(More Sharable)



SHAREABLE



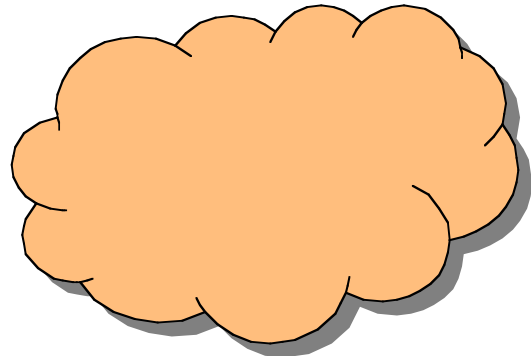
<https://creativecommons.org/licenses/by-nc-sa/4.0/>

© Copyright 2018 by Peter Allen Slide 6 75

Getting data into the Cloud

Problems with forklifting

1. no basis for decisions made
2. no inclusion of architecture/engineering concepts
3. no idea that these concepts are missing from the process
4. 80% of organizational data is ROT
5. all I/O can be chargeable



Transform

Less
Cleaner
More shareable
... data



<https://creativecommons.org/licenses/by-nc-sa/4.0/>

© Copyright 2018 by Peter Allen Slide 6 76

Effective Cloud Transformation



- Transformation into cloud computing cannot be done in a manner that benefits organizations unless data is re-architected – formally with two goals:
 - Maximizing effective, organization-wide data sharing; and
 - Minimizing organizational data ROT.
- Resulting data volume reduction should be 1/5 what is currently is
 - A significant economic motivator.
- All existing organizations have data collections that possess unique strengths and weaknesses
 - Strengths that should be leveraged
 - Weaknesses must be addressed
- Neither of these can be accomplished without formal data re-architecting prior to cloud loading.
- There are very few who work in the area for a living but my teams have achieved some remarkable successes.



Fixing Data in the Cloud Using A Glovebox



The Cloud as a Data Quality Tool

Continuous Improvement
Data Baselineing
Statistical Data Control
Cost of Quality Model
Empowerment

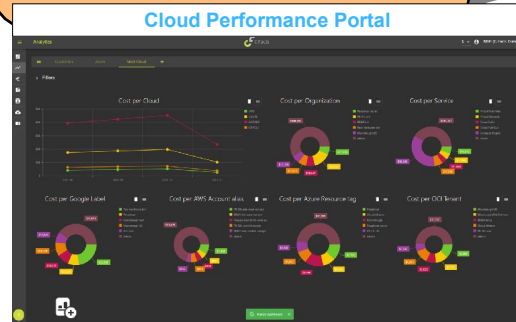
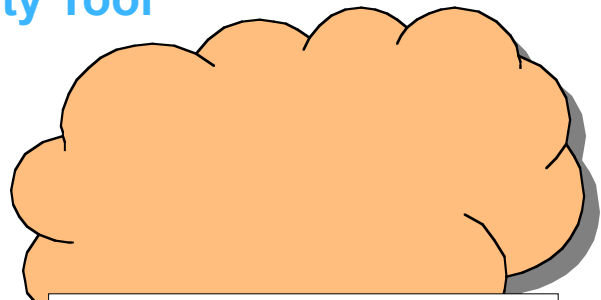
Quality

Data Reduction
Pattern Analysis
Mathematical Analysis
Schema Validation

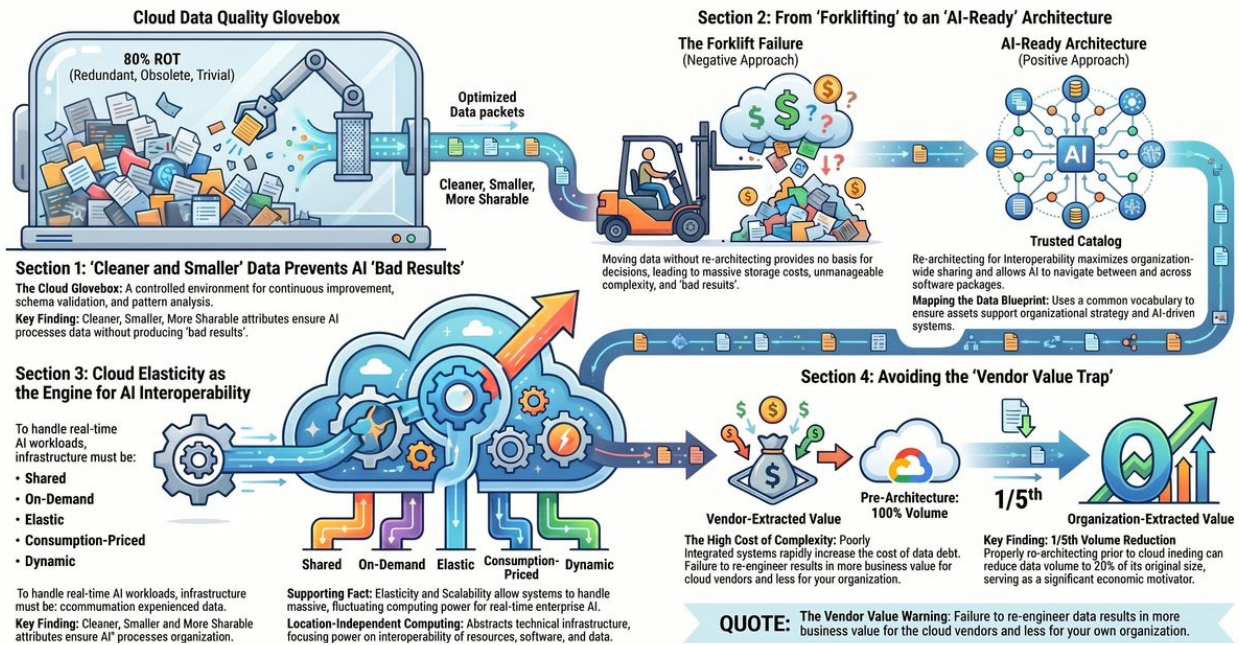
Data Analysis

Technology

Reusability
Logic & Logic Programming
Relational DB Technologies
Data Migration Technologies
Statistical Programming Languages



Powering AI-Driven Cloud Interoperability





Section 1: 'Cleaner and Smaller' Data Prevents AI 'Bad Results'

The Cloud Glovebox: A controlled environment for continuous improvement, schema validation, and pattern analysis.

Key Finding: Cleaner, Smaller, More Sharable attributes ensure AI processes data without producing 'bad results'.



Section 2: From 'Forklifting' to an 'AI-Ready' Architecture

The Forklift Failure
(Negative Approach)



Moving data without re-architecting provides no basis for decisions, leading to massive storage costs, unmanageable complexity, and 'bad results'.

AI-Ready Architecture
(Positive Approach)



Re-architecting for Interoperability maximizes organization-wide sharing and allows AI to navigate between and across software packages.

Mapping the Data Blueprint: Uses a common vocabulary to ensure assets support organizational strategy and AI-driven systems.



Powering AI-Driven Cloud Interoperability



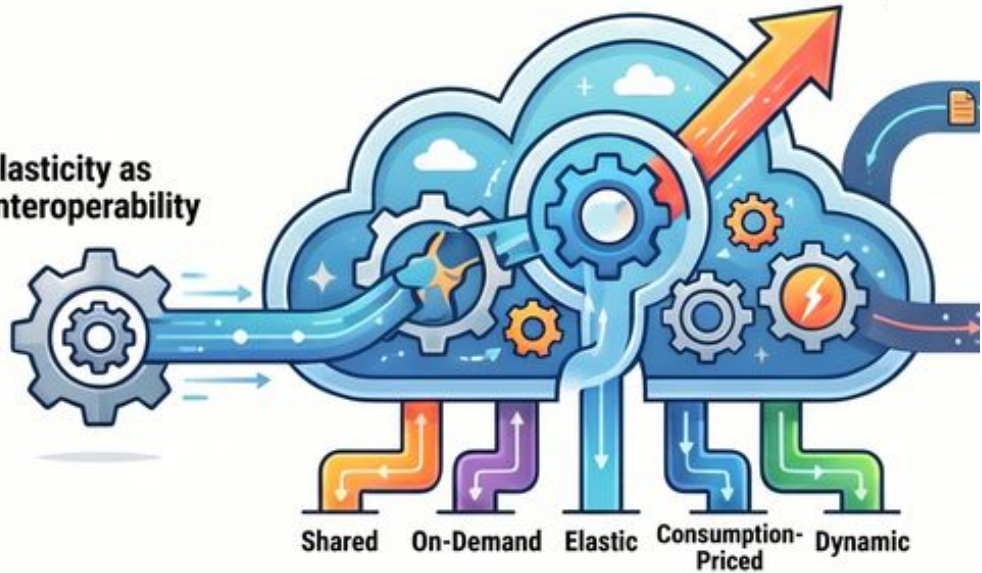
Section 3: Cloud Elasticity as the Engine for AI Interoperability

To handle real-time AI workloads, infrastructure must be:

- Shared
- On-Demand
- Elastic
- Consumption-Priced
- Dynamic

To handle real-time AI workloads, infrastructure must be: communication experienced data.

Key Finding: Cleaner, Smaller and More Sharable attributes ensure AI processes organization.



Supporting Fact: Elasticity and Scalability allow systems to handle massive, fluctuating computing power for real-time enterprise AI.

Location-Independent Computing: Abstracts technical infrastructure, focusing power on interoperability of resources, software, and data.



Powering AI-Driven Cloud Interoperability



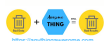
Section 4: Avoiding the 'Vendor Value Trap'



The High Cost of Complexity: Poorly Integrated systems rapidly increase the cost of data debt. Failure to re-engineer results in more business value for cloud vendors and less for your organization.

Key Finding: 1/5th Volume Reduction
Properly re-architecting prior to cloud ineding can reduce data volume to 20% of its original size, serving as a significant economic motivator.

QUOTE: The Vendor Value Warning: Failure to re-engineer data results in more business value for the cloud vendors and less for your own organization.



Program Overview

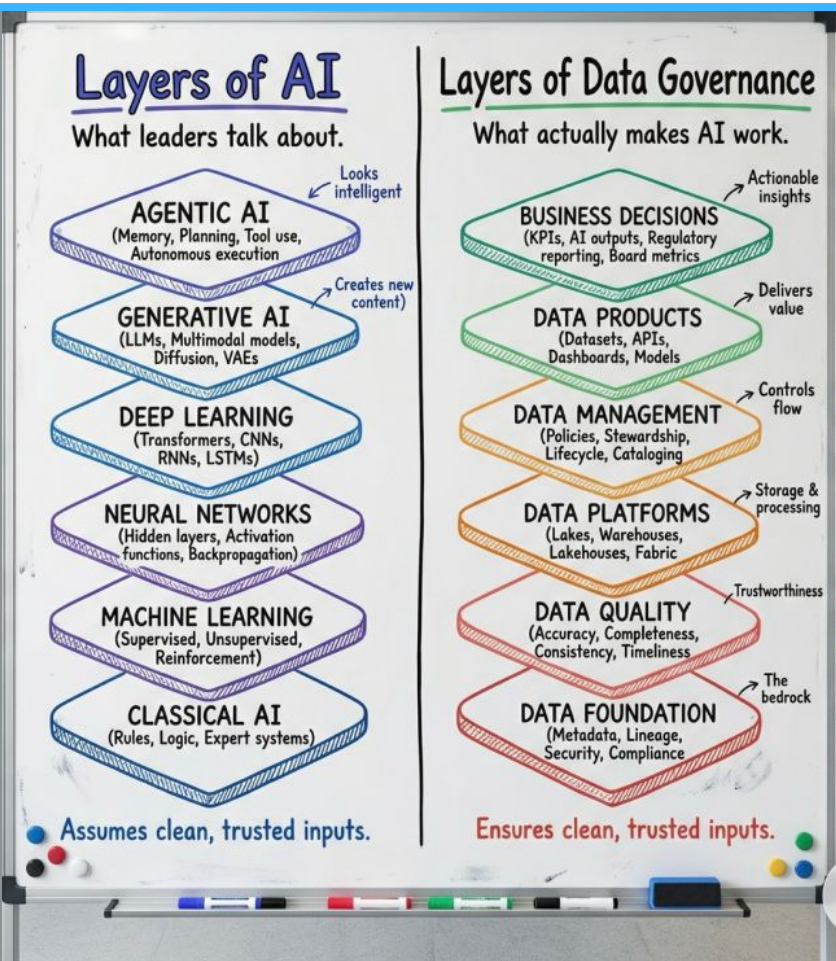
Data Integration Demystified: Enabling AI-Driven Interoperability

- Integration Goals/Challenges
- AI-based approach-augmentation
- Metadata
 - Replace the concept of technical detail with that of a foundational language (common vocabulary) that provides context and clarity for your data and AI
- Cloud
 - Before taking full advantage of the scalability, agility, and cost-effectiveness of cloud services, cloud data should be architected
- Data Warehousing
 - Needs a new layer to prepare for AI
- Take aways, references and Q&A



The key here is to position the data warehouse not just as a reporting tool, but as the structured, high-quality repository that AI algorithms require to function across the enterprise.

Layers of AI and Data Governance

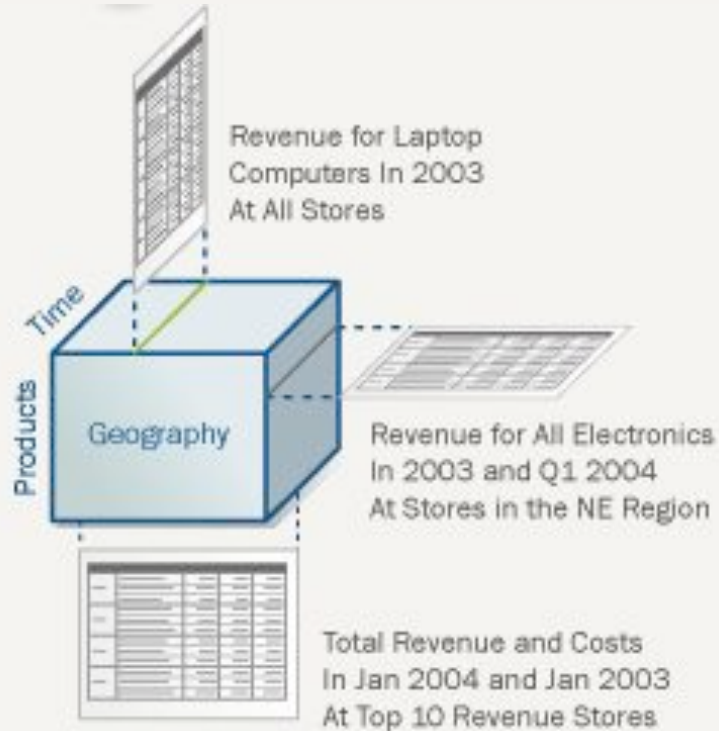


https://www.linkedin.com/posts/john-wernfeldt-82894b58-1frankly-i-keep-hearing-leaders-talk-about-activity-742921662251602912-PCab7?utm_medium=ios_app&utm_source=social_share_send&utm_campaign=messages

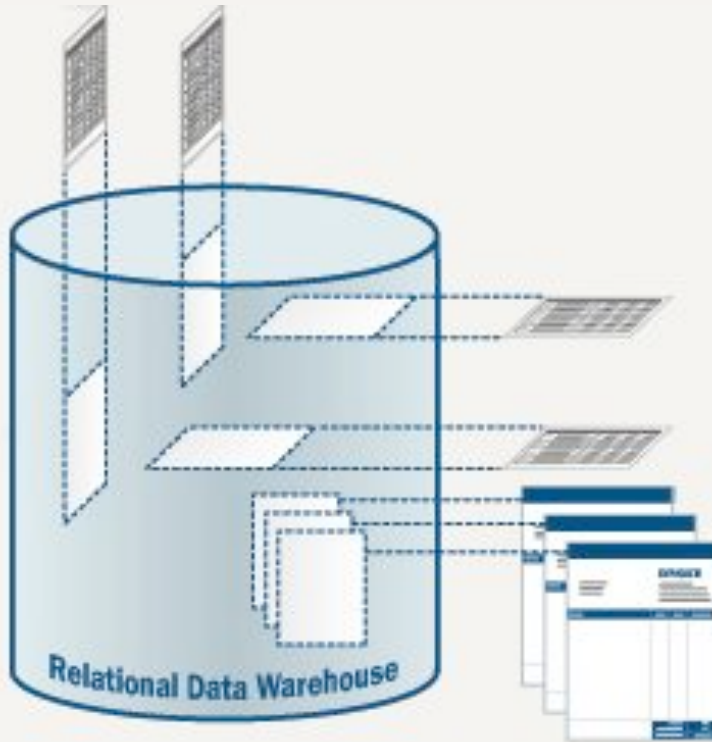


Basic Data Warehouse Analysis

- Emphasis on the cube
- Permits different users to "slice and dice" subsets of data
- Viewing from different perspectives



Warehouse Analysis

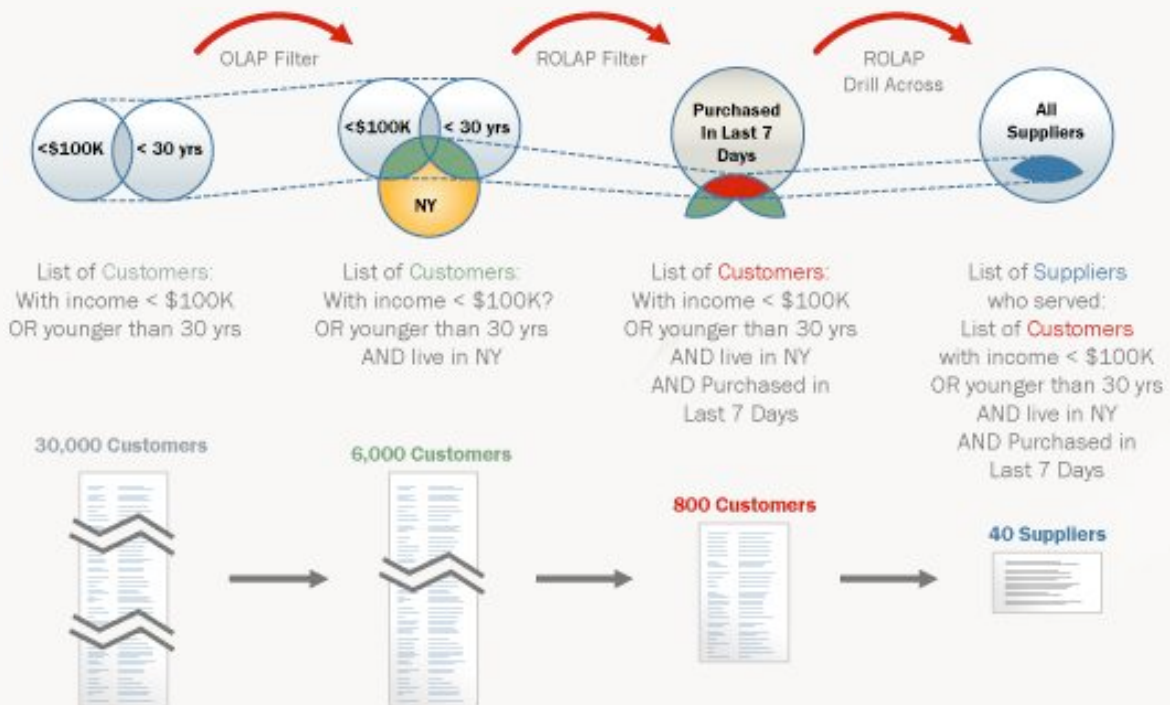


- Users can "drill" anywhere
- Entire collection is accessible
- Summaries to transaction-level detail



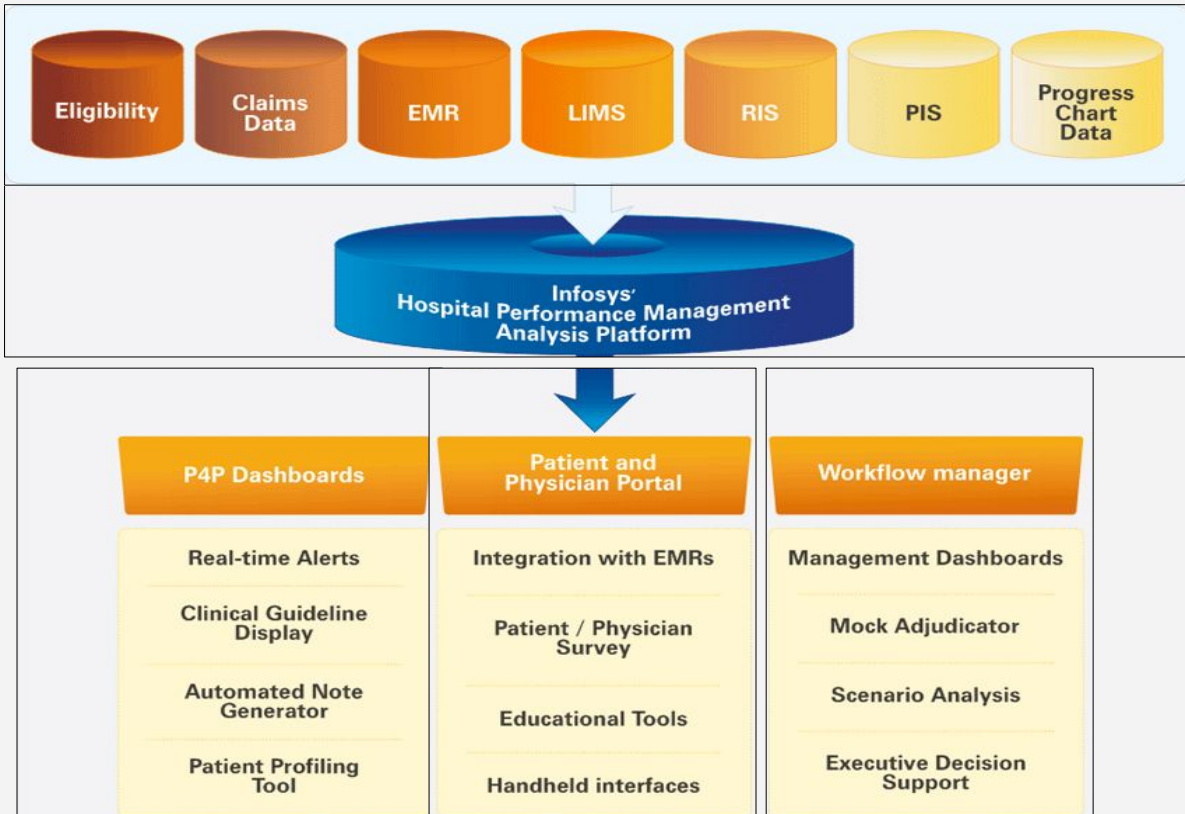
from MicroStrategy, Better Business Decisions Every Day: *Integrating Business Reporting & Analysis*

Example: Set Analysis



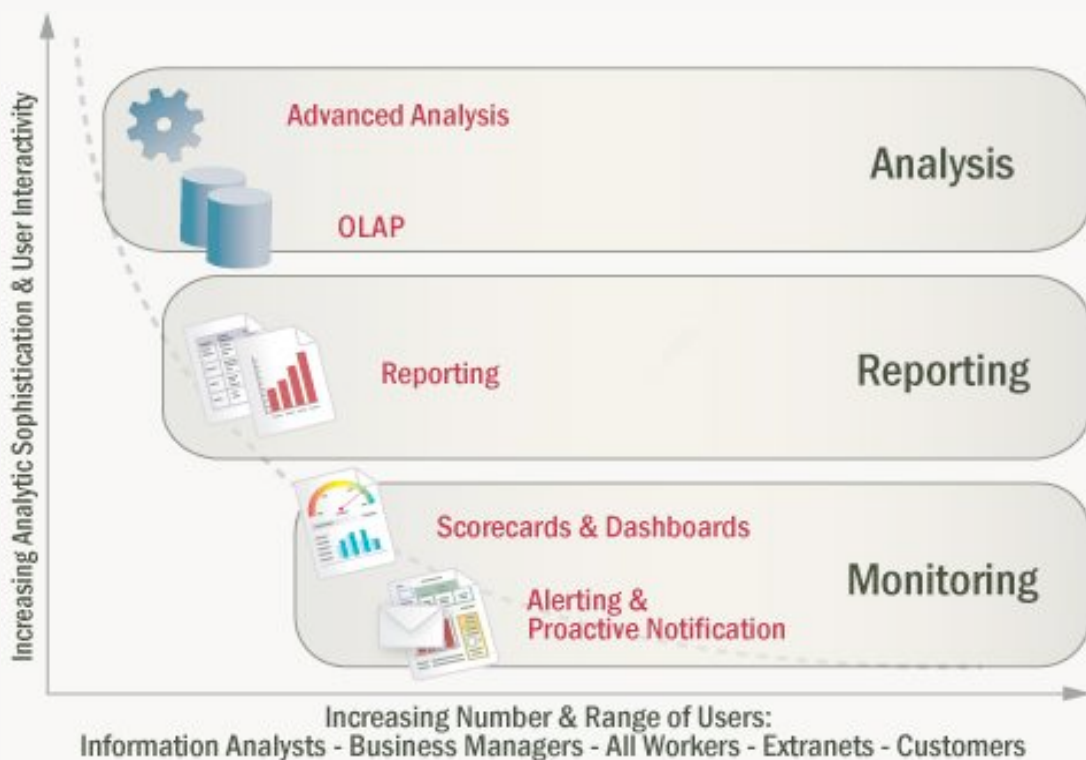
from MicroStrategy, Better Business Decisions Every Day: *Integrating Business Reporting & Analysis*

Data Warehousing



Courtesy of: <http://www.infosys.com/industries/healthcare/industryofferings/Pages/healthcare-data-warehousing.aspx>

Styles of Business Intelligence



from MicroStrategy, Better Business Decisions Every Day: *Integrating Business Reporting & Analysis*

Health Care Provider Data Warehouse

- 1.8 million members
- 1.4 million providers
- 800,000 providers no key
- 29% prov_ssn \neq 9 digits
- 2.2% prov_number = 9 digits (required)
- 1 User



,335	\$	4,812
,280	\$	55,445
,130	\$	(118,059)
,016	\$	(18,016)
.238	\$	(70,786)
4.39	\$	(12.90)
per Visit		

"I can take a roomful of MBAs and accomplish this analysis faster!"

Indiana Jones: Raiders Of The Lost Ark



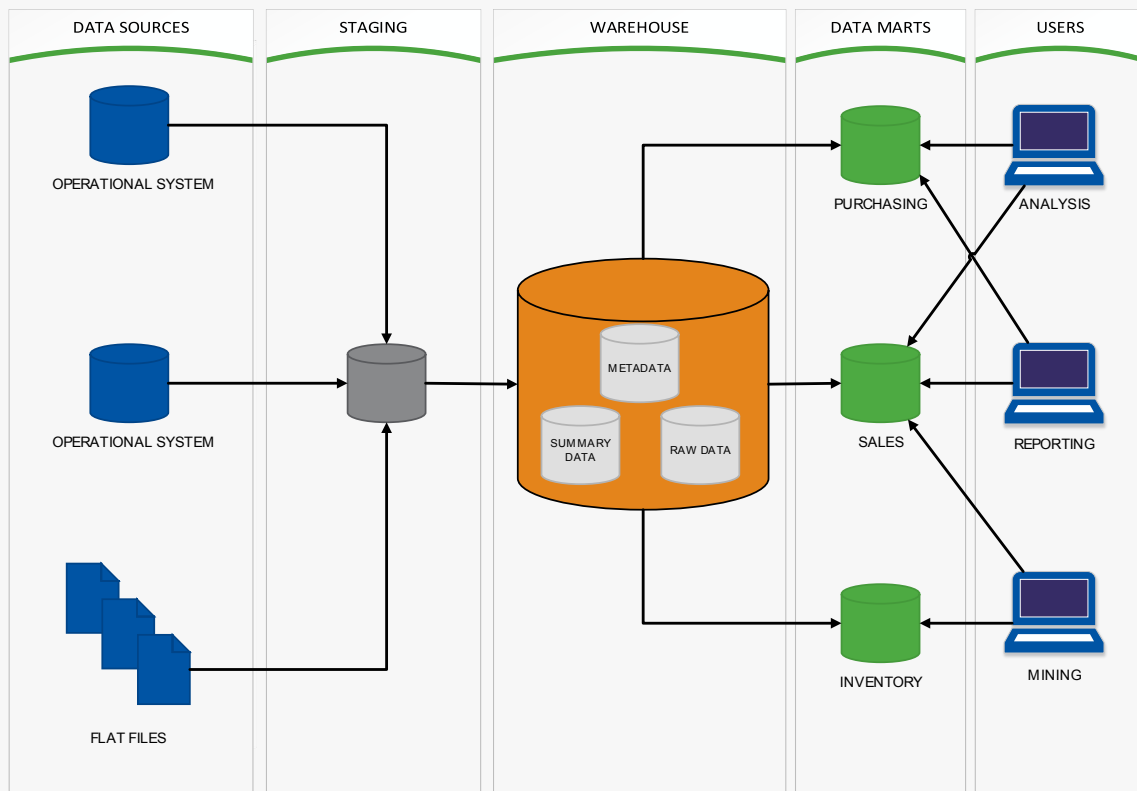
Top Causes of Data Warehouse Failure

- Poor Quality Data
 - Many more values of code than allowable
- Incorrectly Structured Data
 - Providing the correct answer to the wrong question
- Bad Warehouse Design
 - Overly complex

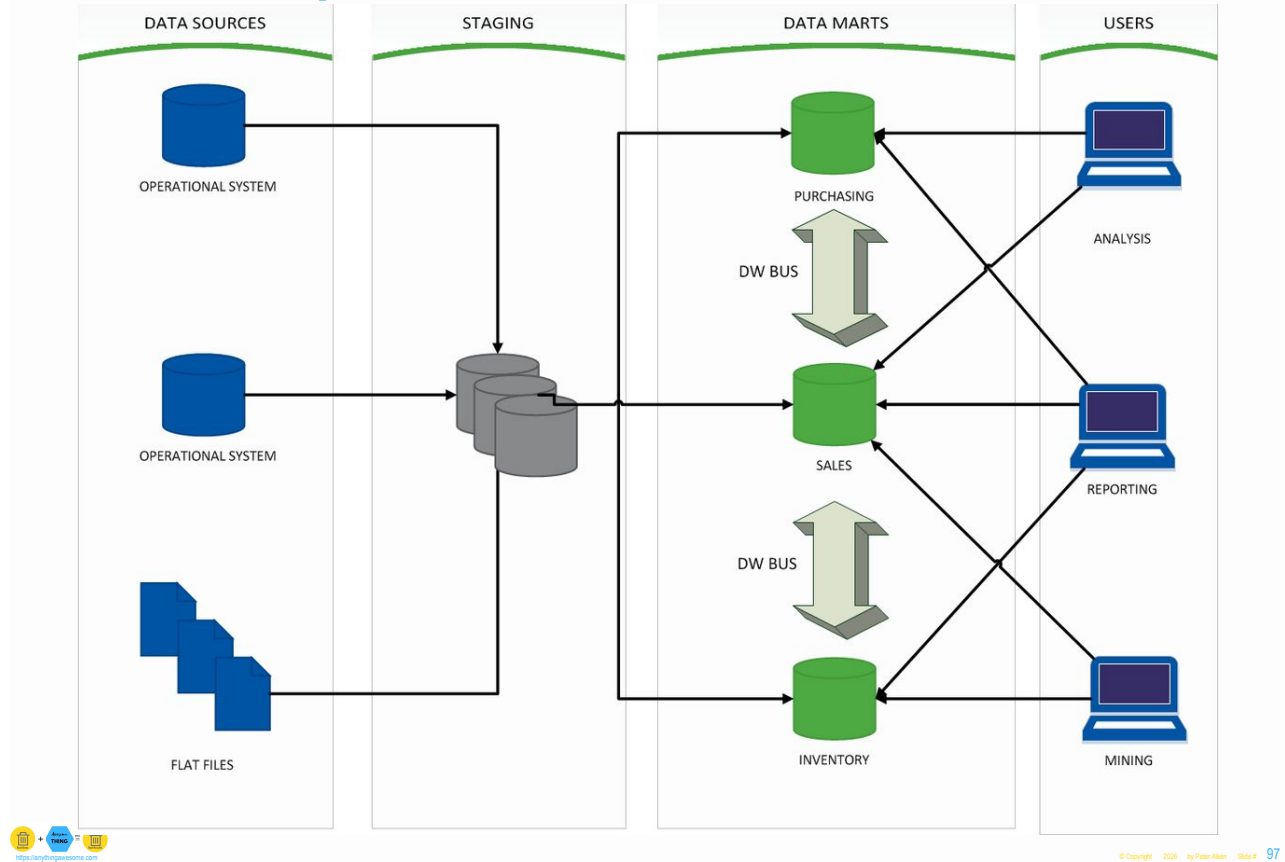


from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International

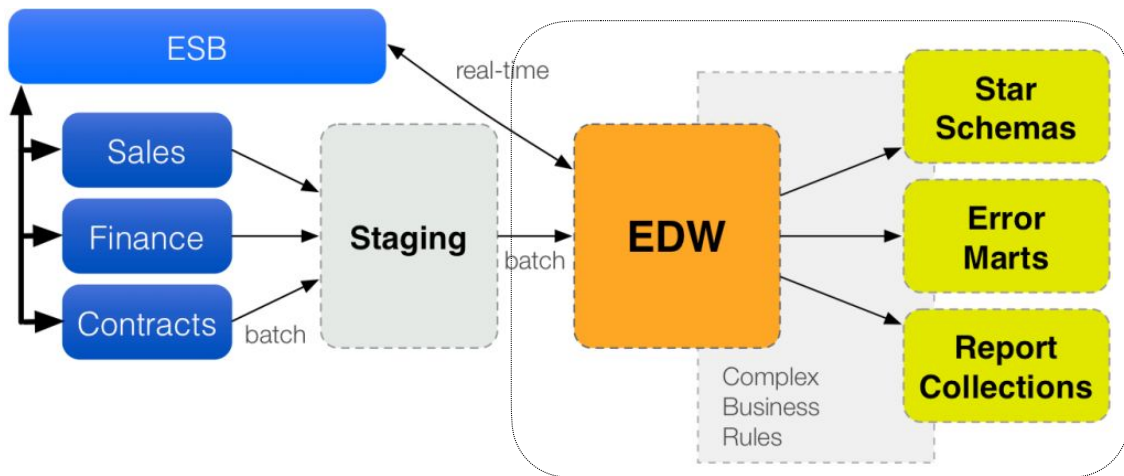
Inmon Implementation/3NF

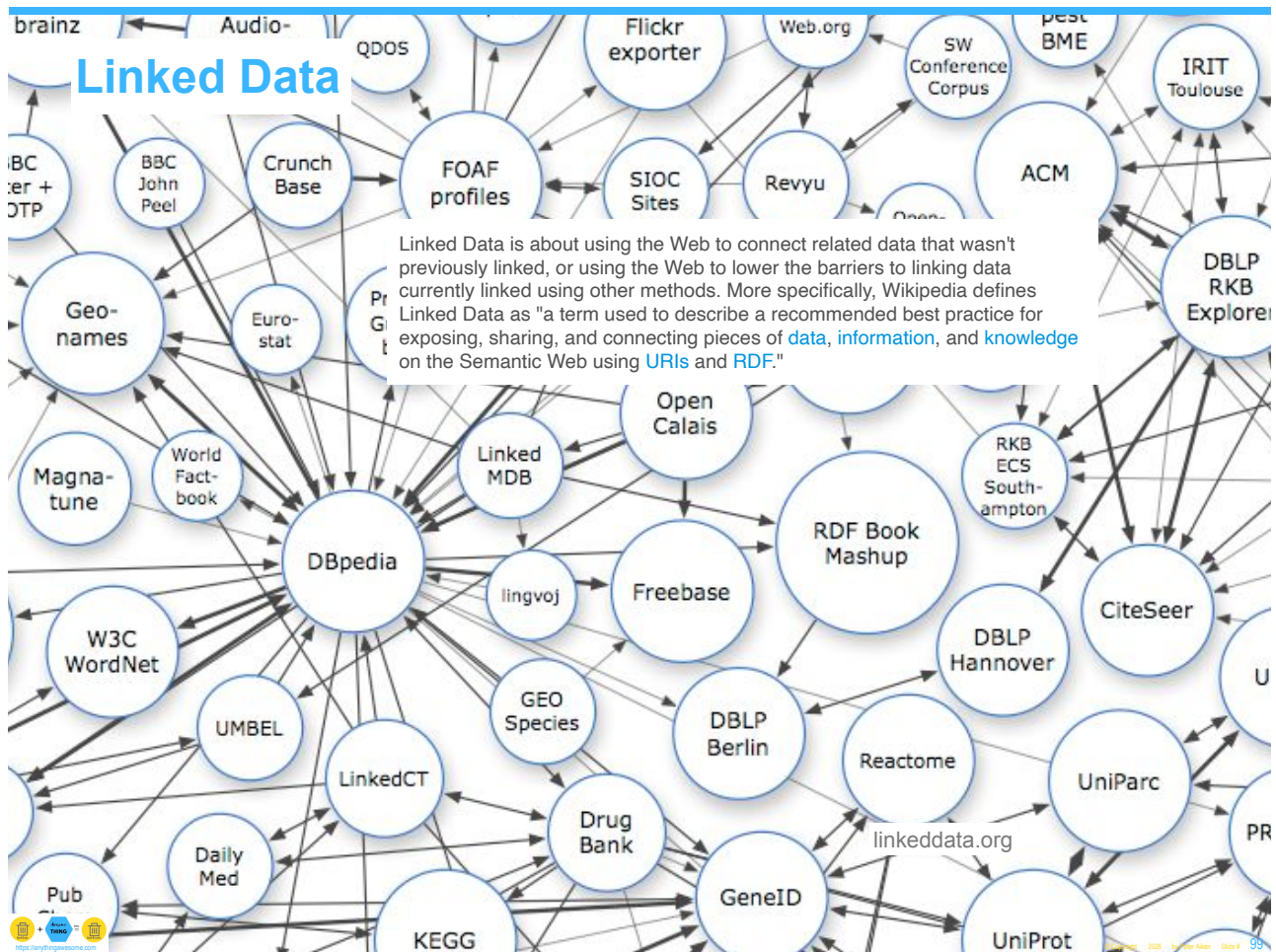


Kimball Implementation/Dimensional

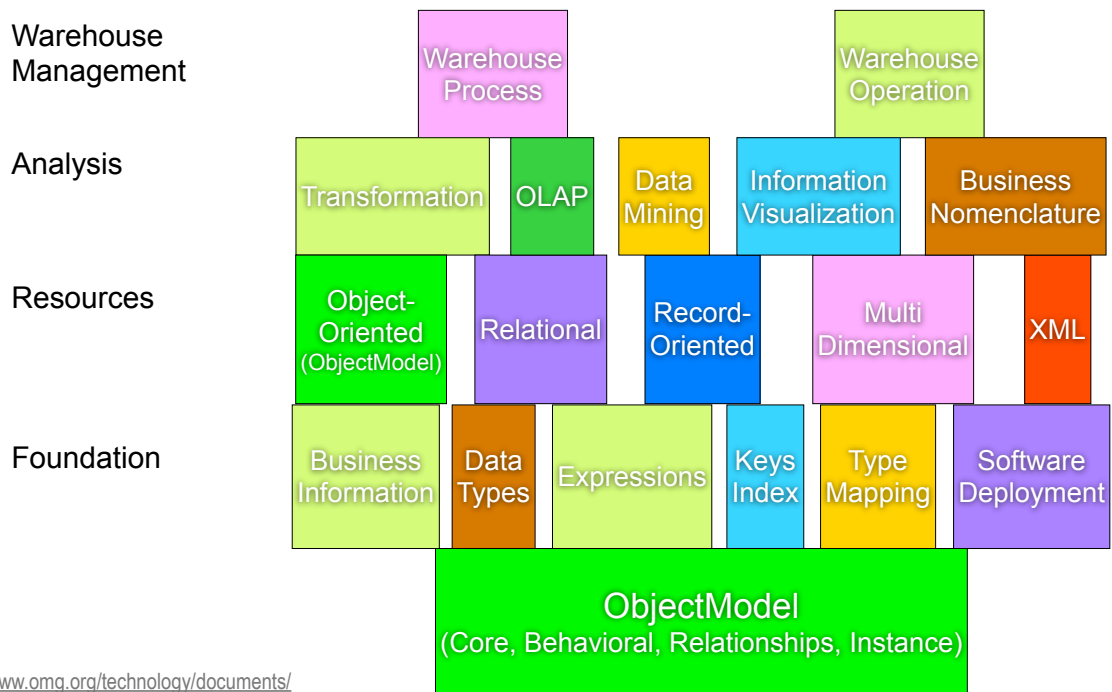


Data Vault Implementation





Overview of CWM Metamodel



http://www.omg.org/technology/documents/modeling_spec_catalog.htm

Powering AI-Driven Warehousing Interoperability



The key here is to position the data warehouse not just as a reporting tool, but as the structured, high-quality repository that AI algorithms require to function across the enterprise.



FRAGMENTED DATA SILOS

Structured Subjects & Transactions (Non-Volatile)

Successful AI integration focuses on structured data subjects and transactions that are not volatile and are rarely restructured for analysis.



Trusted Catalog (Metadata Repository)

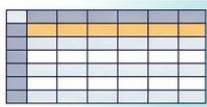
Establishing a common vocabulary and a metadata repository ensures data assets are arranged to support organizational strategy.



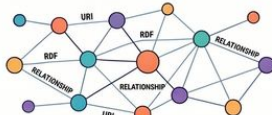
AI-Ready Transaction Data

AI requires a copy of integration data specifically structured for query and analysis, ensuring the "Single Source of Truth" necessary to avoid bad results.

From Relational Tables to 'Linked Data'



Relational Tables



LINKED DATA (SEMANTIC WEB MAP)



The Semantic Web Map

Linked Data uses the Web to connect previously isolated data points, creating a navigable map that AI can follow to discover relationships.



Standardizing with URIs and RDF

Using URIs (Uniform Resource Identifiers) and RDF (Resource Description Framework) provides a recommended best practice for sharing and connecting knowledge.



Lowering Barriers to Integration

By exposing data on the Semantic Web, organizations lower the technical barriers to linking complex, multi-source information.

Warehouse Failures are AI Failures



The Danger of Missing Keys
Data quality issues, such as having 800,000 providers with no identifying key, will paralyze AI and prevent interoperability.



Top Causes of System Failure
Poor quality data (excessive code values), incorrectly structured data, and overly complex designs are the primary drivers of warehouse and AI failure.



"Correct Answer, Wrong Question"
Even structured data fails if the design provides a technically accurate answer to a question that does not serve the business need.

Business Value Must Precede Technical Design



Strategy Over Software
AI in warehousing most target measurable business improvements and direct support of organizational strategy before technical implementation begins.



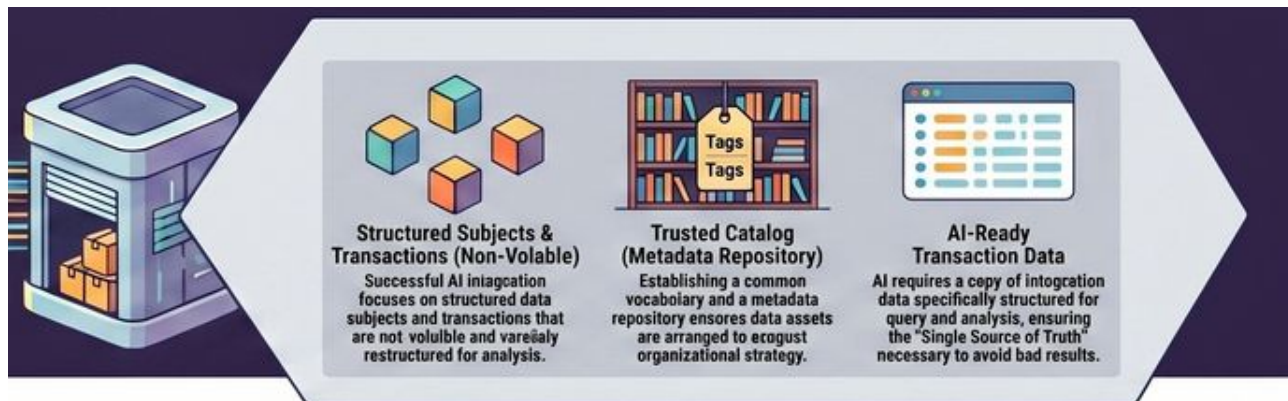
Productive Trade-off Discussions
Achieving the best results requires high-level discussions regarding "buy versus build" and the capabilities required by knowledge workers.



"I can take a roomful of MBAs..."
Technical complexity is no substitute for business-driven analysis, technical design must always follow the definition of business value.



Powering AI-Driven Warehousing Interoperability



Structured Subjects & Transactions (Non-Volatile)

Successful AI integration focuses on structured data subjects and transactions that are not volatile and are rarely restructured for analysis.



Trusted Catalog (Metadata Repository)

Establishing a common vocabulary and a metadata repository ensures data assets are arranged to support organizational strategy.

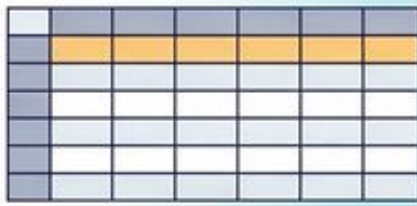


AI-Ready Transaction Data

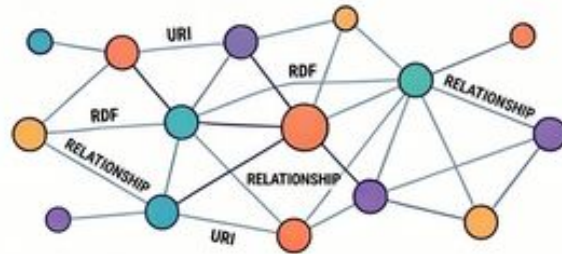
AI requires a copy of integration data specifically structured for query and analysis, ensuring the "Single Source of Truth" necessary to avoid bad results.



Powering AI-Driven Warehousing Interoperability



Relational Tables



LINKED DATA (SEMANTIC WEB MAP)



Powering AI-Driven Warehousing Interoperability



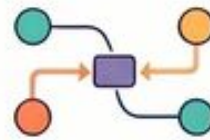
The Semantic Web Map

Linked Data uses the Web to connect previously isolated data points, creating a navigable map that AI can follow to discover relationships.



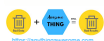
Standardizing with URIs and RDF

Using URIs (Uniform Resource Identifiers) and RDF (Resource Description Framework) provides a recommended best practice for sharing and connecting knowledge.



Lowering Barriers to Integration

By exposing data on the Semantic Web, organizations lower the technical barriers to linking complex, multi-source information.



Program Overview

Data Integration Demystified: Enabling AI-Driven Interoperability

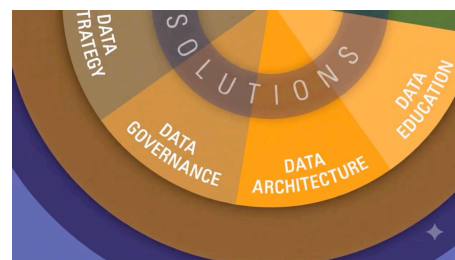
- Integration Goals/Challenges
- AI-based approach-augmentation
- Metadata
 - Replace the concept of technical detail with that of a foundational language (common vocabulary) that provides context and clarity for your data and AI
- Cloud
 - Before taking full advantage of the scalability, agility, and cost-effectiveness of cloud services, cloud data should be architected
- Data Warehousing
 - Needs a new layer to prepare for AI
- Take aways, references and Q&A



Questions to Ask




- Are you ready for a _____?
 - Foundational Practices
- Is the business environment constantly evolving?
- Will you get it right the first time?
- Do you have an agreed-upon ways to maintain an enterprise-wide vocabulary?
- Is your system intended to be an auditable system of record?
- Do we understand the requisite extract, transform and load data transformations?
- How fast do you need results?
- Do we have forecasts of the performance of inserts vs reads



- Metadata
 - Metadata unlocks the value of data, and therefore requires management attention
 - Metadata is the language of data governance
 - Metadata defines the essence of integration challenges
- Cloud
 - Data governance, architecture, quality, development maturity are necessary but insufficient prerequisites to successful data cloud implementation
 - A variety of cloud options will influence cloud and data architectures in general
 - You must understand your architecture and strategy in order to evaluate the options
 - Data must be reengineered to be: less; better quality; more shareable
 - Failure to do these will result in more business value for the cloud vendors/ service providers and less for your organization
- Warehousing
 - Business value must precede technical design
 - Other non-traditional alternatives are available



The Art of Data: Managing Provenance, Market, and Institutional Value




Metadata: The Language of Art Governance

- **Unlocking Data Value for Management**
Provenance research transforms "dead" documentation into nuanced stories about global history, requiring management to prioritize identifying descendants and stakeholders for meaningful engagement.
- **The Language of Data Governance**
Accession ledgers and databases act as the "governance" of an institution, encoding legacies of power that must be critically analyzed to ensure ethical, anticolonial museum practices.
- **Defining Integration Challenges**
Documenting art requires managing "unverifiable stories" and gaps in ownership histories through transparent narrative formatting rather than technical shorthand.

- **Metadata**
 - Metadata unlocks the value of data, and therefore requires management attention
 - Metadata is the language of data governance
 - Metadata defines the essence of integration challenges

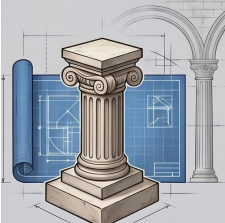
Cloud: Architecture of the Global Art Market

- **Prerequisites for Successful Implementation**
Success in the global market requires analyzing the "Great Wealth Transfer," as \$83 trillion passes to a new generation of women and younger collectors with evolving motivations.
- **Understanding Market Architecture & Strategy**
The 2025 market grew to \$59.6 billion, necessitating a strategy that balances a rebound in high-end auction sales with the \$9.2 billion online art sector.
- **Reengineering Data for Quality and Sharing**
Artists and galleries use technology, such as limited-edition digital works and computer art, to control sales and ensure better quality control over shareable media.
- **The Risks of Failure and Exclusion**
A failed "Cloud" strategy risks art becoming an exclusive upper-class activity where the status marker is "understanding" the work, accessible only to an educated elite.



Global Art Cloud Market Share (2025)

United States	44%	(Up 5% \$26B)
United Kingdom	18%	(Up 2% \$10.5B)
China	14%	(Up 1% \$8.5B)
France	8%	(Up 9% \$4.5B)



Warehousing: Institutional Value & Foundations

Business Value Must Precede Technical Design
Institutional "warehousing" must address the ethical question of "to whom heritage belongs?" as a primary value before refining technical methods like accession ledgers or receipts.

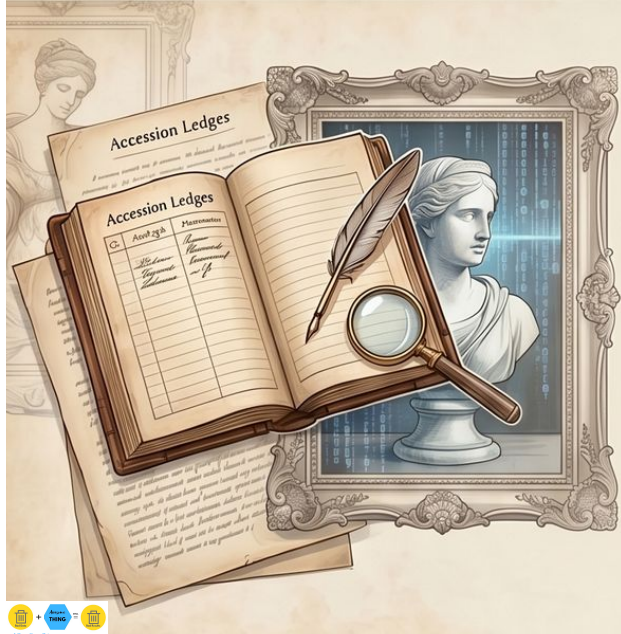
Non-Traditional Warehousing Alternatives
Beyond physical museums, the art world utilizes virtual museums, artist-run spaces, sculpture trails, and alternative exhibition sites to preserve and display creative works.

- **Cloud**
 - Data governance, architecture, quality, development maturity are necessary but insufficient prerequisites to successful data cloud implementation
 - A variety of cloud options will influence cloud and data architectures in general
 - You must understand your architecture and strategy in order to evaluate the options
 - Data must be reengineered to be: less; better quality; more shareable
 - Failure to do these will result in more business value for the cloud vendors/service providers and less for your organization
- **Warehousing**
 - Business value must precede technical design
 - Other non-traditional alternatives are available



The Art of Data:

Managing Provenance, Market, and Institutional Value

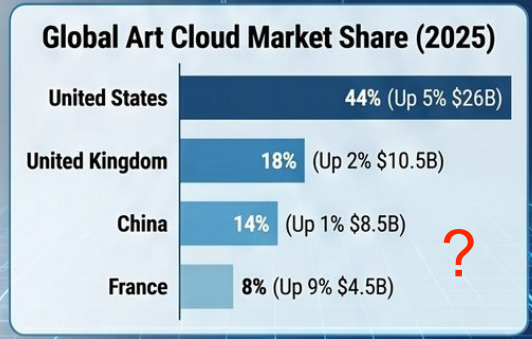


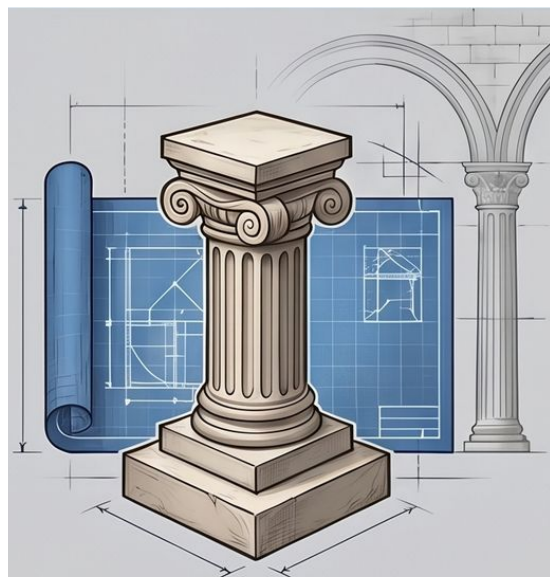
Metadata: The Language of Art Governance

- **Unlocking Data Value for Management**
Provenance research transforms a “dead” documentation into nuanced stories about global history, requiring management to prioritize identifying descendants and stakeholders for meaningful engagement.
- **The Language of Data Governance**
Accession ledgers and databases act as the “governance” of an institution, encoding legacies of power that must be critically analyzed to ensure ethical, anticolonial museum practices.
- **Defining Integration Challenges**
Documenting art requires managing “unverifiable stories” and gaps in ownership histories through transparent narrative formatting rather than technical shorthand.

Cloud: Architecture of the Global Art Market

- **Prerequisites for Successful Implementation**
? Success in the global market requires analyzing the “Great Wealth Transfer,” as \$83 trillion passes to a new generation of women and younger collectors with evolving motivations.
- **Understanding Market Architecture & Strategy**
? The 2025 market grew to \$59.6 billion, necessitating a strategy that balances a rebound in high-end auction sales with the \$9.2 billion online art sector.
- **Reengineering Data for Quality and Sharing**
? Artists and galleries use technology, such as limited-edition digital works and computer art, to control sales and ensure better quality control over shareable media.
- **The Risks of Failure and Exclusion**
? A failed “Cloud” strategy risks art becoming an exclusive upper-class activity where the status marker is “understanding” the work, accessible only to an educated elite.





Warehousing: Institutional Value & Foundations

Business Value Must Precede Technical Design
 Institutional “warehousing” must address the ethical question of “to whom heritage belongs” as a primary value before refining technical methods like accession ledgers or receipts.

Non-Traditional Warehousing Alternatives
 Beyond physical museums, the art world utilizes virtual museums, artist-run spaces, sculpture trails, and alternative exhibition sites to preserve and display creative works.

- Best Results Achieved with:
 - Trade-off productive discussions
 - Buying versus build discussions are similarly productive
 - Direct support of strategy
 - Measurable business improvements
 - Support for knowledge worker capabilities



© Copyright 2026 by Peter Aiken Slide # 111

Upcoming Events Time: 19:00 UTC (2:00 PM NYC) | Presented by: Peter Aiken, PhD

AI-Powered Data Modeling Essentials

14 July 2026

Mastering Reference Data: An AI Essential for Reliable Business Information

11 August 2026



Data Warehousing Reimagined: Fueling AI-Powered BI at Scale

8 September 2026

Brought to you by:

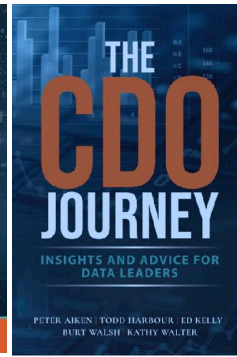
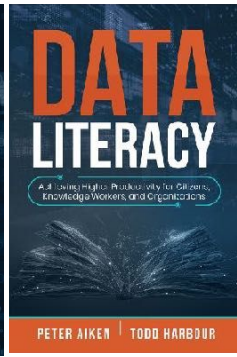
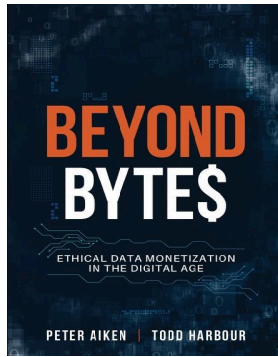


[Clicking any webinar title will link directly to the registration page]

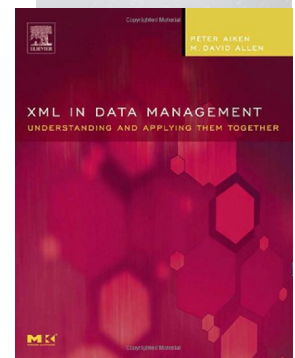
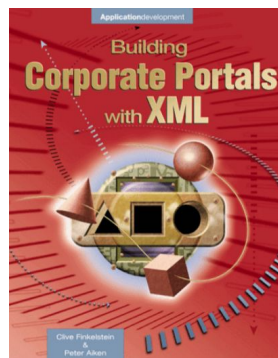
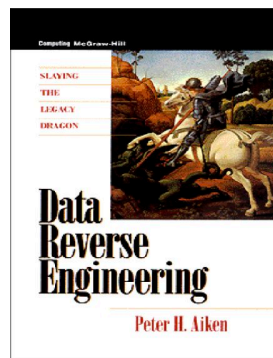
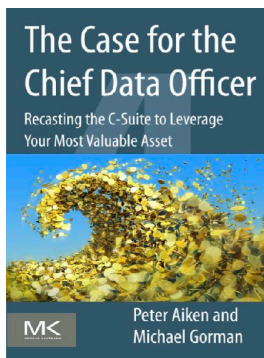
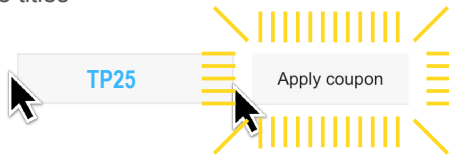


© Copyright 2026 by Peter Aiken Slide # 112

Event Pricing on Peter's Books



- 25% off directly from the publisher on some titles
- @ <https://anythingawesome.com/books>
- Enter the code "TP25" at the Technics bookstore checkout where it says to "Apply Coupon"



Critical Design Review?

Independent Verification & Validation



Mentoring?

Executive Data Literacy Training?

Collaboration?



Peter.Aiken@AnythingAwesome.com +1.804.382.5957

Reverse Engineering Expertise?

Hiring Assistance?

Thank You!

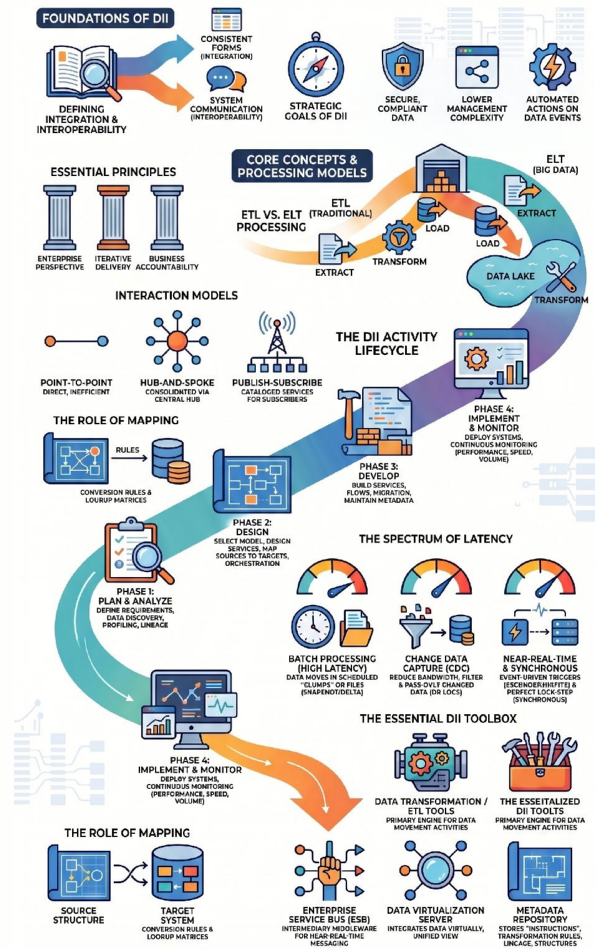
Use your data more strategically?

Tool/automation evaluation?

Book a call with Peter to discuss anything - <https://anythingawesome.com/OfficeHours.html>



References



Data Integration and Interoperability (DII)

DAMA DM BoK® Chapter 8

Definition and Goals: Data Integration and Interoperability (DII) describes the processes related to the movement and consolidation of data within and between data stores, applications, and organizations. The primary goals of DII are to:

- Provide data securely and in regulatory compliance, in the format and timeframe needed.
- Lower costs and complexity by developing shared models and interfaces.
- Identify meaningful events to automatically trigger alerts and actions.
- Support business intelligence (BI), analytics, Master Data Management (MDM), and operational efficiency.

Essential Concepts

- **ETL and ELT:** Central to DII is the Extract, Transform, and Load (ETL) process, which selects data from a source, makes it compatible with a target structure, and physically stores it. Alternatively, ELT (Extract, Load, and Transform) loads raw data directly into the target system before transforming it, which is especially common in Big Data environments.
- **Latency:** Data movement happens at different speeds, ranging from high-latency batch processing (e.g., nightly scheduled runs) to low-latency/streaming (real-time continuous flow). Other methods include Change Data Capture (CDC) to pull only recent changes, and asynchronous or synchronous real-time updates.
- **Interaction Models:** Systems can interact via point-to-point connections (direct but highly complex to scale), hub-and-spoke models (using a central data hub to minimize interfaces), or publish-subscribe models (where systems push data out and consumers pull it in).



- **Architecture Concepts:** Mature DII leverages architectures like Enterprise Service Buses (ESB) for near-real-time messaging, Service-Oriented Architecture (SOA) for independent data services, and Complex Event Processing (CEP) to analyze data streams and predict behaviors.

Key Activities The DII lifecycle is divided into four main phases:

- **Plan & Analyze:** Define integration and lifecycle requirements, discover data sources, document data lineage, profile data for quality and structure, and harvest business rules.
- **Design DII Solutions:** Formulate the architecture, map source data to target data, and design the orchestration (the pattern and timing of data flows).
- **Develop DII Solutions:** Build data services, orchestrate data flows, develop data migration approaches, implement complex event processing, and capture DII metadata (like transformation rules and lineage).
- **Implement and Monitor:** Activate services and monitor processing for performance issues or automated event triggers.
- **Tools and Governance** Executing DII requires specialized tools such as data transformation engines (ETL tools), data virtualization servers, Enterprise Service Buses, business rules engines, and metadata repositories to track transformation rules and lineage.

DII governance ensures that integration is driven by business rules rather than purely technical approaches. It involves establishing Data Sharing Agreements (MOUs stipulating acceptable use and access), strictly managing data lineage for impact analysis, and tracking key metrics like data availability, volume, speed, and solution costs.



Data is Your Most Valuable Asset: Here's How to Manage It

Treat data with the same rigor as financial and physical assets to derive ongoing value.

Guiding Principles of Data Management

Data is a Unique Enterprise Asset

Unlike other assets, it's not consumed when used and its value can grow over time.



Metadata is Essential for Management

To manage data, you need data about your data (Metadata) to provide context and meaning.



Management Spans the Full Data Lifecycle



Data must be managed at every stage:
Plan & Design / Create / Store / Use / Dispose.

Key Challenges in Data Management

Quantifying Data's Value is Difficult

Unlike financial assets, there are no standard methods for expressing data's economic value.



Poor Data Quality is Costly

Poor quality data cost the US an estimated \$3.1 Trillion in 2016 and erodes trust.



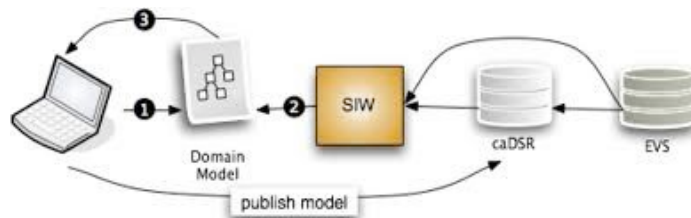
Overcoming Data Silos is Critical

Data originates in different departments, creating inconsistent and conflicting views of the same business concepts.

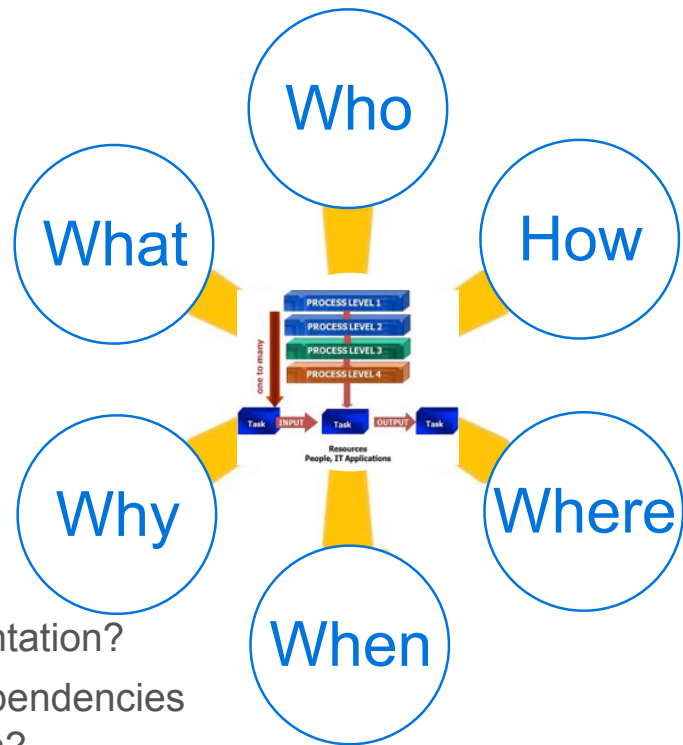


Types of Metadata: Process Metadata

- Process Metadata is...
 - Data that defines and describes the characteristics of other system elements, e.g. processes, business rules, programs, jobs, tools, etc.
- Examples of Process metadata:
 - Data stores and data involved
 - Government/regulatory bodies
 - Organization owners and stakeholders
 - Process dependencies and decomposition
 - Process feedback loop and documentation
 - Process name



Types of Metadata: Process Metadata



Who: Created the documentation?

What: Are the important dependencies among the processes?

How: Do the business processes interact with each other?



Source: Brad Melton

© Copyright 2008 by Peter Allen Slide # 119

Types of Metadata: Business Metadata

- Business Metadata describe to the end user what data are available, what they mean and how to retrieve them.

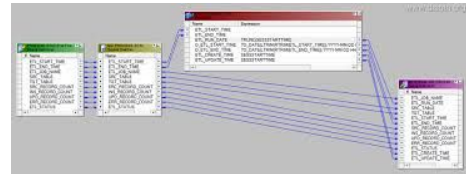


- Included are:
 - Business names and definitions of subject and concept areas, entities, attributes
 - Attribute data types and other attribute properties
 - Range descriptions, calculations, algorithms and business rules
 - Valid domain values and their definitions



Types of Metadata: Technical & Operational Metadata

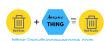
- Technical and operational metadata provides developers and technical users with information about their systems
- Technical metadata includes...
 - Physical database table and column names, column properties, other properties, other database object properties and database storage
- Operational metadata is targeted at IT operations users' needs, including...
 - Information about data movement, source and target systems, batch programs, job frequency, schedule anomalies, recovery and backup information, archive rules and usage
- Examples of Technical & Operational metadata:
 - Audit controls and balancing information
 - Data archiving and retention rules
 - Encoding/reference table conversions
 - History of extracts and results



from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International

Types of Metadata: Data Stewardship

- Data stewardship Metadata is about...
 - Data stewards, stewardship processes, and responsibility assignments
- Data stewards...
 - Assure that data and Metadata are accurate, with high quality across the enterprise.
 - Establish and monitor data sharing.
- Examples of Data stewardship metadata:
 - Business drivers/goals
 - Data CRUD rules
 - Data definitions – business and technical
 - Data requirements ownership
 - Data sharing rules and agreements/contracts
 - Data stewards, roles and responsibilities



from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International

Types of Metadata: Provenance

- Provenance:
 - the history of ownership of a valued object or work of art or literature" [Merriam Webster]
 - For each datum, this is the description of:
 - Its source (system or person or department),
 - Any derivation used, and
 - The date it was created.
 - Examples of Data Provenance:
 - The programs or processes by which it was created
 - Its owner
 - The steward responsible for its quality
 - Other roles and responsibilities
 - Rules for sharing it

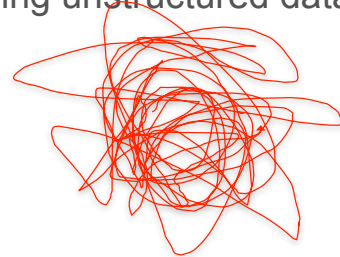


from The DAMA Guide to the Data Management Body of Knowledge © 2009 by DAMA International

© Copyright 2009 by Peter Allen Slide # 123

Metadata for Semistructured Data

- Unstructured data
 - Any data that is not in a database or data file, including documents or other media data
- Metadata describes both structured and unstructured data
- Metadata for unstructured data exists in many formats, responding to a variety of different requirements
- Examples of Metadata repositories describing unstructured data:
 - Content management applications
 - University websites
 - Company intranet sites
 - Data archives
 - Electronic journals collections
 - Community resource lists
- Common method for classifying Metadata in unstructured sources is to describe them as descriptive metadata, structural metadata, or administrative metadata



from The DAMA Guide to the Data Management Body of Knowledge © 2009 by DAMA International 24

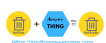
Metadata for Unstructured Data: Examples

- (Preferred) Non-tabular Data
- Examples of descriptive metadata:
 - Catalog information
 - Thesauri keyword terms
- Examples of structural metadata
 - Dublin Core
 - Field structures
 - Format (audio/visual, booklet)
 - Thesauri keyword labels
 - XML schemas
- Examples of administrative metadata
 - Source(s)
 - Integration/update schedule
 - Access rights
 - Page relationships (e.g. site navigational design)



Sources of Metadata

- Primary Sources:
 - Virtually anything named in an organization
- Secondary sources:
 - Other Metadata repositories, accessed using bridge software
 - CASE tools, ETL tools
- Many data management tools create and use repositories for their own use.



References

Adamson, Christopher. Mastering Data Warehouse Aggregates: Solutions for Star Schema Performance. John Wiley & Sons, 2006. ISBN 0-471-77709-9. 345 pages.

Adamson, Christopher and Michael Venerable. Data Warehouse Design Solutions. John Wiley & Sons, 1998. ISBN 0-471-25195-X. 544 pages.

Adelman, Sid and Larissa T. Moss. Data Warehouse Project Management. Addison-Wesley Professional, 2000. ISBN 0-201-61635-1. 448 pages.

Adelman, Sid and others. Impossible Data Warehouse Situations: Solutions from the Experts. Addison-Wesley, 2002. ISBN 0-201-76033-9. 432 pages.

Brackett, Michael. The Data Warehouse Challenge: Taming Data Chaos. New York: John Wiley & Sons, 1996. ISBN 0-471-12744-2. 579 pages.

Caserta, Joe and Ralph Kimball. The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming and Delivering Data. John Wiley & Sons, 2004. ISBN 0-764-56757-8. 525 pages.



References

Correy, Michael J. and Michael Abby. Oracle Data Warehousing: A Practical Guide to Successful Data Warehouse Analysis, Build and Roll-Out. TATA McGraw-Hill, 1997. ISBN 0-074-63069-5.

Covey, Stephen R. The 7 Habits of Highly Effective People. Free Press, 2004. ISBN 0743269519. 384 Pages.

Dyche, Jill. E-Data: Turning Data Into Information With Data Warehousing. Addison-Wesley, 2000. ISBN 0-201-65780-5. 384 pages.

Gill, Harjinder S. and Prekash C. Rao. The Official Guide To Data Warehousing. Que, 1996. ISBN 0-789-70714-4. 382 pages.

Hackney, Douglas. Understanding and Implementing Successful Data Marts. Addison Wesley, 1997. ISBN 0-201-18380-3. 464 pages.

Imhoff, Claudia, Nicholas Gallemmo and Jonathan G. Geiger. Mastering Data Warehouse Design: Relational and Dimensional Techniques. John Wiley & Sons, 2003. ISBN 0-471-32421-3. 456 pages.

Imhoff, Claudia, Lisa Loftis and Jonathan G. Geiger. Building the Customer-Centric Enterprise: Data Warehousing Techniques for Supporting Customer Relationship Management. John Wiley & Sons, 2001. ISBN 0-471-31981-3. 512 pages.

Inmon, W. H. Building the Data Warehouse, 4th Edition. John Wiley & Sons, 2005. ISBN 0-764-50944-5. 543 pages.

Inmon, W. H. Building the Operational Data Store, 2nd edition. John Wiley & Sons, 1999. ISBN 0-471-32888-X. 336 pages.

Inmon, W. H., Claudia Imhoff and Ryan Sousa. The Corporate Information Factory, 2nd edition. John Wiley & Sons, 2000. ISBN 0-471-39961-2. 400 pages.

Inmon, W. H. and Richard D. Hackathorn. Using the Data Warehouse. Wiley-QED, 1994. ISBN 0-471-05966-8. 305 pages.

Inmon, William H., John A. Zachman and Jonathan G. Geiger. Data Stores, Data Warehousing and the Zachman Framework. McGraw-Hill, 1997. ISBN 0-070-31429-2. 358 pages.

Kimball, Ralph and Margy Ross. The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling, 2nd edition. New York: John Wiley & Sons, 2002. ISBN 0-471-20024-7. 464 pages.

Kimball, Ralph, Laura Reeves, Margy Ross and Warren Thornwaite. The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing and Deploying Data Warehouses. John Wiley & Sons, 1998. ISBN 0-471-25547-5. 800 pages.

Kimball, Ralph and Richard Merz. The Data Warehouse Toolkit: Building the Web-Enabled Data Warehouse. John Wiley & Sons, 2000. ISBN 0-471-37680-9. 416 pages.

Mattison, Rob. Web Warehousing & Knowledge Management. McGraw Hill, 1999. ISBN 0-070-41103-4. 576 pages.

Morris, Henry. Analytic Applications and Business Performance Management. DM Review Magazine, March, 1999. www.dmreview.com. Note: www.dmreview.com is now www.information-management.com.

Moss, Larissa T. and Shaku Atré. Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications. Addison-Wesley, 2003. ISBN 0-201-78420-3. 576 pages.

Poe, Vidette, Patricia Klauer and Stephen Brobst. Building A Data Warehouse for Decision Support, 2nd edition. Prentice-Hall, 1987. ISBN 0-137-69638-6. 285 pages.

Ponniiah, Paulraj. Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals. John Wiley & Sons - Interscience, 2001. ISBN 0-471-41254-6. 528 pages.

Westerman, Paul. Data Warehousing: Using the Wal-Mart Model. Morgan Kaufman, 2000. ISBN 155860684X. 297 pages.

9.5.2 Business Intelligence

Biore, Mike. Business Intelligence for the Enterprise. IBM Press, 2003. ISBN 0-131-41303-1. 240 pages.

Eckerson, Wayne W. Performance Dashboards: MEasuring, Monitoring, and Managing Your Business. Wiley, 2005. ISBN-10: 0471724173. 320 pages.

Bischoff, Joyce and Ted Alexander. Data Warehouse: Practical Advice from the Experts. Prentice-Hall, 1997. ISBN 0-135-77370-9. 428 pages.

Howson, Cindi. "The Business Intelligence Market". <http://www.biscurecard.com/>. Requires annual subscription to this website.

Malik, Shadan. Enterprise Dashboards: Design and Best Practices for IT. Wiley, 2005. ISBN 0471738069. 240 pages.

Moss, Larissa T., and Shaku Atré. Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications. Addison-Wesley, 2003. ISBN 0-201-78420-3. 576 pages.

Vitt, Elizabeth, Michael Luckevich and Stacia Misner. Business Intelligence. Microsoft Press, 2008. ISBN 073562660X. 220 pages.

9.5.3 Data Mining

Cabena, Peter, Hadjioan, Stadler, Verhees and Zanasi. Discovering Data Mining: From Concept to Implementation. Prentice Hall, 1997. ISBN-10: 0137439806



References & Recommended Reading

11.4.1 General Reading

Brathwaite, Ken S. Analysis, Design, and Implementation of Data Dictionaries. McGraw-Hill Inc., 1988. ISBN 0-07-007248-5. 214 pages.

Collier, Ken. Executive Report, Business Intelligence Advisory Service, *Finding the Value in Metadata Management* (Vol. 4, No. 1), 2004. Available only to Cutter Consortium Clients, <http://www.cutter.com/bia/fulltext/reports/2004/01/index.html>.

Hay, David C. Data Model Patterns: A Metadata Map. Morgan Kaufmann, 2006. ISBN 0-120-88798-3. 432 pages.

Hillmann, Diane I. and Elaine L. Westbrooks, editors. Metadata in Practice. American Library Association, 2004. ISBN 0-838-90882-9. 285 pages.

Inmon, William H., Bonnie O'Neil and Lowell Fryman. Business Metadata: Capturing Enterprise Knowledge. 2008. Morgan Kaufmann ISBN 978-0-12-373726-7. 314 pages.

Marco, David, Building and Managing the Meta Data Repository: A Full Life-Cycle Guide. John Wiley & Sons, 2000. ISBN 0-471-35523-2. 416 pages.

Marco, David and Michael Jennings. Universal Meta Data Models. John Wiley & Sons, 2004. ISBN 0-471-08177-9. 478 pages.

Poole, John, Dan Change, Douglas Tolbert and David Mellor. Common Warehouse Metamodel: An Introduction to the Standard for Data Warehouse Integration. John Wiley & Sons, 2001. ISBN 0-471-20052-2. 208 pages.



from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International²⁹

References, cont'd

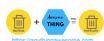
Poole, John, Dan Change, Douglas Tolbert and David Mellor. Common Warehouse Metamodel Developer's Guide. John Wiley & Sons, 2003. ISBN 0-471-20243-6. 704 pages.

Ross, Ronald. Data Dictionaries And Data Administration: Concepts and Practices for Data Resource Management. New York: AMACOM Books, 1981. ISN 0-814-45596-4. 454 pages.

Tannenbaum, Adrienne. Implementing a Corporate Repository, John Wiley & Sons, 1994. ISBN 0-471-58537-8. 441 pages.

Tannenbaum, Adrienne. Metadata Solutions: Using Metamodels, Repositories, XML, And Enterprise Portals to Generate Information on Demand. Addison Wesley, 2001. ISBN 0-201-71976-2. 528 pages.

Wertz, Charles J. The Data Dictionary: Concepts and Uses, 2nd edition. John Wiley & Sons, 1993. ISBN 0-471-60308-2. 390 pages.



from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International³⁰

References, cont'd

11.4.2 Meta-data in Library Science

Baca, Murtha, editor. Introduction to Metadata: Pathways to Digital Information. Getty Information Institute, 2000. ISBN 0-892-36533-1. 48 pages.

Hillmann, Diane I., and Elaine L. Westbrooks. Metadata in Practice. American Library Association, 2004. ISBN 0-838-90882-9. 285 pages.

Karpuk, Deborah. METADATA: From Resource Discovery to Knowledge Management. Libraries Unlimited, 2007. ISBN 1-591-58070-6. 275 pages.

Liu, Jia. Metadata and Its Applications in the Digital Library. Libraries Unlimited, 2007. ISBN 1-291-58306-6. 250 pages.

11.4.3 Geospatial Meta-data Standards

<http://www.fgdc.gov/metadata/geospatial-metadata-standards>.



from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International³¹

References, cont'd

11.4.4 ISO Meta-data Standards

ISO Standards Handbook 10, Data Processing—Vocabulary, 1982.

ISO 704:1987, Principles and methods of terminology.

ISO 1087, Terminology—Vocabulary.

ISO 2382-4:1987, Information processing systems—Vocabulary part 4.

ISO/IEC 10241:1992, International Terminology Standards—Preparation and layout.

FCD 11179-2, Information technology—Specification and standardization of data elements - Part 2: Classification for data elements.

ISO/IEC 11179-3:1994, Information technology—Specification and standardization of data elements - Part 3: Basic attributes of data elements.

ISO/IEC 11179-4:1995, Information technology—Specification and standardization of data elements - Part 4: Rules and guidelines for the formulation of data definitions.

ISO/IEC 11179-5:1995, Information technology—Specification and standardization of data elements - Part 5: Naming and identification principles for data elements.

ISO/IEC 11179-6:1997, Information technology—Specification and standardization of data elements - Part 6: Registration of data elements.



from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International³²

References

Adamson, Christopher. Mastering Data Warehouse Aggregates: Solutions for Star Schema Performance. John Wiley & Sons, 2006. ISBN 0-471-77709-9. 345 pages.

Adamson, Christopher and Michael Venerable. Data Warehouse Design Solutions. John Wiley & Sons, 1998. ISBN 0-471-25195-X. 544 pages.

Adelman, Sid and Larissa T. Moss. Data Warehouse Project Management. Addison-Wesley Professional, 2000. ISBN 0-201-61635-1. 448 pages.

Adelman, Sid and others. Impossible Data Warehouse Situations: Solutions from the Experts. Addison-Wesley, 2002. ISBN 0-201-76033-9. 432 pages.

Brackett, Michael. The Data Warehouse Challenge: Taming Data Chaos. New York: John Wiley & Sons, 1996. ISBN 0-471-12744-2. 579 pages.

Caserta, Joe and Ralph Kimball. The Data Warehouse ETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming and Delivering Data. John Wiley & Sons, 2004. ISBN 0-764-56757-8. 525 pages.



References

Correy, Michael J. and Michael Abby. Oracle Data Warehousing: A Practical Guide to Successful Data Warehouse Analysis, Build and Roll-Out. TATA McGraw-Hill, 1997. ISBN 0-074-63069-5.

Covey, Stephen R. The 7 Habits of Highly Effective People. Free Press, 2004. ISBN 0743269519. 384 Pages.

Dyche, Jill. E-Data: Turning Data Into Information With Data Warehousing. Addison-Wesley, 2000. ISBN 0-201-65780-5. 384 pages.

Gill, Harjinder S. and Prekash C. Rao. The Official Guide To Data Warehousing. Que, 1996. ISBN 0-789-70714-4. 382 pages.

Hackney, Douglas. Understanding and Implementing Successful Data Marts. Addison Wesley, 1997. ISBN 0-201-18380-3. 464 pages.

Imhoff, Claudia, Nicholas Galleo and Jonathan G. Geiger. Mastering Data Warehouse Design: Relational and Dimensional Techniques. John Wiley & Sons, 2003. ISBN 0-471-32421-3. 456 pages.

Imhoff, Claudia, Lisa Loftis and Jonathan G. Geiger. Building the Customer-Centric Enterprise: Data Warehousing Techniques for Supporting Customer Relationship Management. John Wiley & Sons, 2001. ISBN 0-471-31981-3. 512 pages.

Inmon, W. H. Building the Data Warehouse, 4th Edition. John Wiley & Sons, 2005. ISBN 0-764-50944-5. 543 pages.

Inmon, W. H. Building the Operational Data Store, 2nd edition. John Wiley & Sons, 1999. ISBN 0-471-32888-X. 336 pages.

Inmon, W. H., Claudia Imhoff and Ryan Sousa. The Corporate Information Factory, 2nd edition. John Wiley & Sons, 2000. ISBN 0-471-39961-2. 400 pages.

Inmon, W. H. and Richard D. Hackathorn. Using the Data Warehouse. Wiley-QED, 1994. ISBN 0-471-05966-8. 305 pages.

Inmon, William H., John A. Zachman and Jonathan G. Geiger. Data Stores, Data Warehousing and the Zachman Framework. McGraw-Hill, 1997. ISBN 0-070-31429-2. 358 pages.

Kimball, Ralph and Margy Ross. The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling, 2nd edition. New York: John Wiley & Sons, 2002. ISBN 0-471-20024-7. 464 pages.

Kimball, Ralph, Laura Reeves, Margy Ross and Warren Thornwaite. The Data Warehouse Lifecycle Toolkit: Expert Methods for Designing, Developing and Deploying Data Warehouses. John Wiley & Sons, 1998. ISBN 0-471-25547-5. 800 pages.

Kimball, Ralph and Richard Merz. The Data Warehouse Toolkit: Building the Web-Enabled Data Warehouse. John Wiley & Sons, 2000. ISBN 0-471-37680-9. 416 pages.

Mattison, Rob. Web Warehousing & Knowledge Management. McGraw Hill, 1999. ISBN 0-070-41103-4. 576 pages.

Morris, Henry. Analytic Applications and Business Performance Management. DM Review Magazine, March, 1999. www.dmreview.com. Note: www.dmreview.com is now www.information-management.com.

Moss, Larissa T. and Shaku Atré. Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications. Addison-Wesley, 2003. ISBN 0-201-78420-3. 576 pages.

Poe, Vidette, Patricia Klauer and Stephen Brobst. Building A Data Warehouse for Decision Support, 2nd edition. Prentice-Hall, 1987. ISBN 0-137-69638-6. 285 pages.

Ponniiah, Paulraj. Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals. John Wiley & Sons - Interscience, 2001. ISBN 0-471-41254-6. 528 pages.

Westerman, Paul. Data Warehousing: Using the Wal-Mart Model. Morgan Kaufman, 2000. ISBN 155860684X. 297 pages.

9.5.2 Business Intelligence

Biore, Mike. Business Intelligence for the Enterprise. IBM Press, 2003. ISBN 0-131-41303-1. 240 pages.

Eckerson, Wayne W. Performance Dashboards: MEasuring, Monitoring, and Managing Your Business. Wiley, 2005. ISBN-10: 0471724173. 320 pages.

Bischoff, Joyce and Ted Alexander. Data Warehouse: Practical Advice from the Experts. Prentice-Hall, 1997. ISBN 0-135-77370-9. 428 pages.

Howson, Cindi. "The Business Intelligence Market". <http://www.biscurecard.com/>. Requires annual subscription to this website.

Malik, Shadan. Enterprise Dashboards: Design and Best Practices for IT. Wiley, 2005. ISBN 0471738069. 240 pages.

Moss, Larissa T., and Shaku Atré. Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications. Addison-Wesley, 2003. ISBN 0-201-78420-3. 576 pages.

Vitt, Elizabeth, Michael Luckevich and Stacia Misner. Business Intelligence. Microsoft Press, 2008. ISBN 073562660X. 220 pages.

9.5.3 Data Mining

Cabena, Peter, Hadjioan, Stadler, Verhees and Zanasi. Discovering Data Mining: From Concept to Implementation. Prentice Hall, 1997. ISBN-10: 0137439805



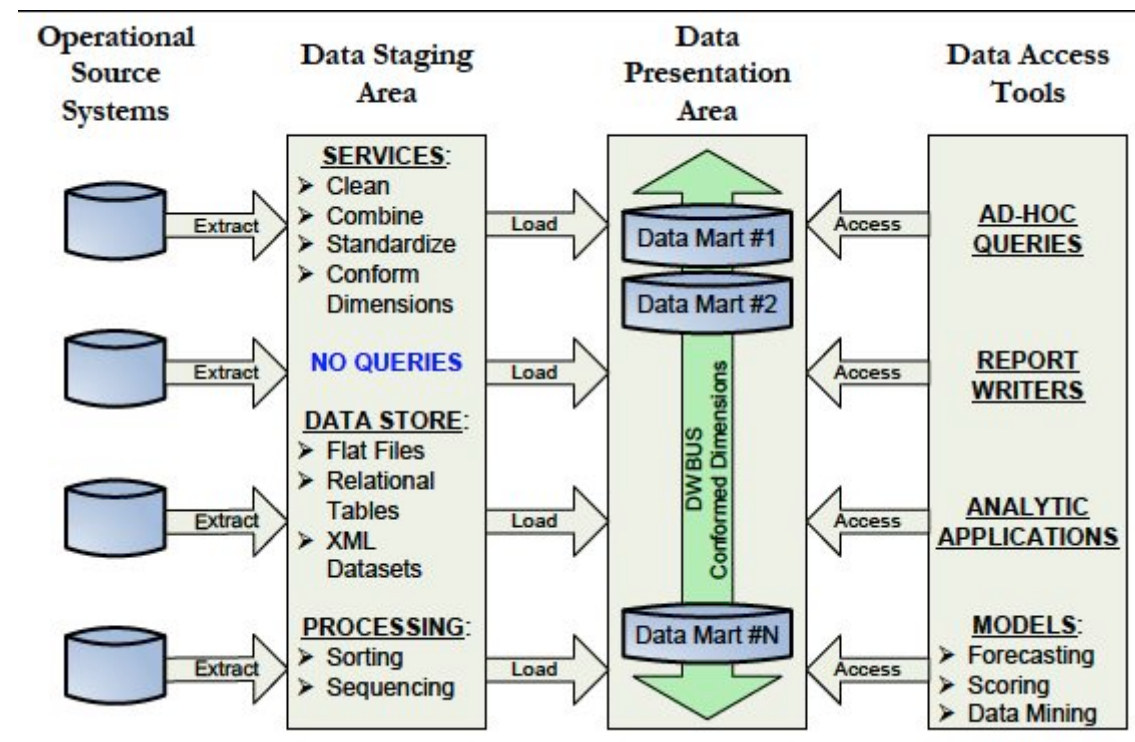
Additional References

- <http://www.information-management.com/infodirect/20050909/1036703-1.html>
- <http://www.agiledata.org/essays/dataWarehousingBestPractices.html>
- http://www.cio.com/article/150450/Five_Key_Business_Intelligence_Trends_You_Need_to_Know?page=2&taxonomyId=3002
- http://www.computerworld.com/s/article/9228736/Business_Intelligence_and_analytics_Conquering_Big_Data?taxonomyId=9
- <http://www.enterpriseirregulars.com/5706/the-top-10-trends-for-2010-in-analytics-business-intelligence-and-performance-management/>
- <http://www.itbusinessedge.com/cm/blogs/vizard/taking-the-analytics-pressure-off-the-data-warehouse/?cs=50698>
- <http://www.informationweek.com/news/software/bi/240001922>

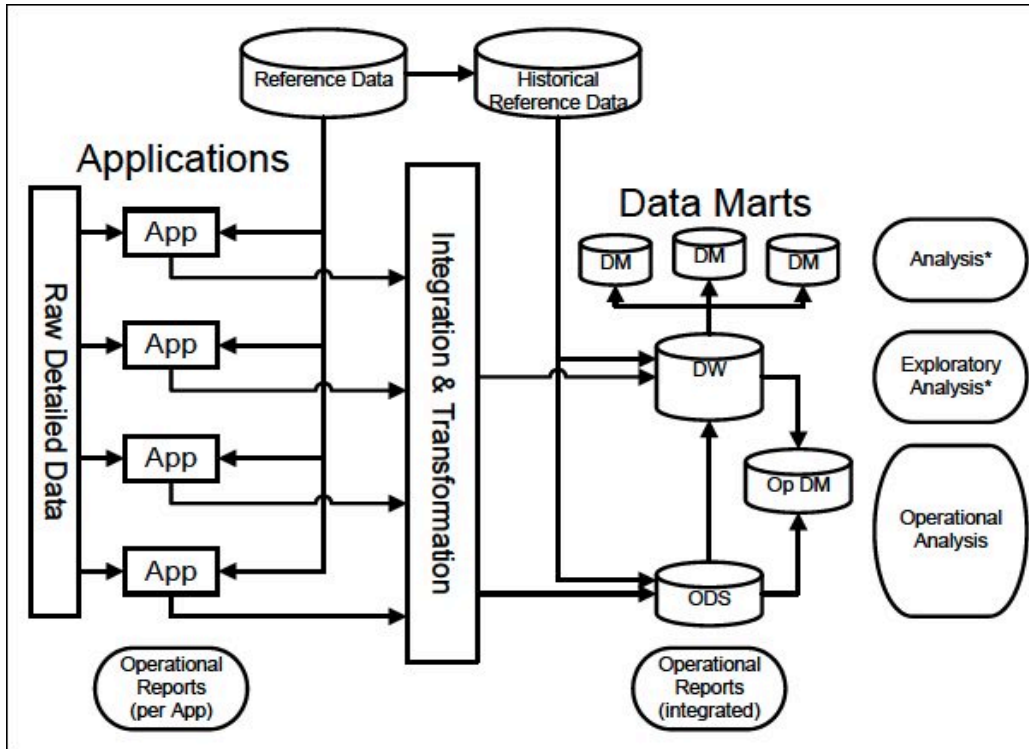


from *The DAMA Guide to the Data Management Body of Knowledge* © 2009 by DAMA International

Kimball's DW Chess Pieces



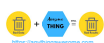
Corporate Information Factory Architecture



Corporate Information Factory Architecture

<p>Raw Detailed data</p>	<p>Operational / Transactional Application data of the enterprise. The raw detailed data provides the source data to be integrated into the Operational Data Store (ODS) and DW components. They can also be in database or other storage or file format.</p>
<p>Integration and Transformation</p>	<p>This layer of the architecture is where the un-integrated data from the various application sources stores is combined / integrated and transformed into the corporate representation in the DW.</p>
<p>Reference Data</p>	<p>Reference data was a precursor to what is currently referred to as Master Data Management. The purpose was to allow common storage and access for important and frequently used common data. Focus and shared understanding on data upstream of the Data Warehouse simplifies the integration task in the DW.</p>
<p>Historical Reference Data</p>	<p>When current valued reference data is necessary for transactional applications, and at the same time it is critical to have accurate integration and presentation of historical data, it is necessary to capture the reference data that was in place at any point in time. For more discussion on reference data, see Chapter 8 Master and Reference Data Management.</p>

from The DAMA Guide to the Data Management Body of Knowledge © 2009 by DAMA International



Corporate Information Factory Architecture

Operational Data Store (ODS)	<p>The focus of data integration is meeting operating and classically operational reporting needs that require data from multiple operational systems. The main distinguishing data characteristics of an ODS compared to a DW include current-valued vs. DW historical data and volatile vs. DW non-volatile data.</p> <p>Note: ODS is an optional portion of the overall CIF architecture, dependent upon specific operational needs, and acknowledged as a component that many businesses omit.</p>
Operational Data Mart (Oper-Mart)	<p>A data mart focuses on tactical decision support. Distinguishing characteristics include current-valued vs. DW historical data, tactical vs. DW strategic analysis, and sourcing of data from an ODS rather than just the DW. The Oper-Mart was a later addition to the CIF architecture.</p>
Data Warehouse (DW)	<p>The DW is a large, comprehensive corporate resource, whose primary purpose is to provide a single integration point for corporate data in order to serve management decision, and strategic analysis and planning.</p> <p>The data flows into a DW from the application systems and ODS, and flows out to the data marts, usually in one direction only. Data that needs correction is rejected, corrected at its source, and re-fed through the system.</p>
Data Marts (DM)	<p>The purpose of the data marts is to provide for DSS / information processing and access that is customized and tailored for the needs of a particular department or common analytic need.</p>

from The DAMA Guide to the Data Management Body of Knowledge © 2009 by DAMA International



Corporate Information Factory Architecture

Applications	Isolated Operational Reports	Limited to data within one application instance
ODS	Integrated Operational Reports	Reports requiring data from multiple source systems. Typically, they have more operational than analytical orientation, with little historical data.
DW	Exploratory Analysis	The complete set of corporate data allows for discovery of new relationships and information. Many BI data mining tools work with flat-file extracts from the DW, which can also offload the processing burden from the DW.
Oper-Mart	Tactical Analytics	Analytic reporting based on current-values with a tactical focus. Dimensional data modeling techniques employed.
Data Mart	Analytics – classical management decision support, and Strategic Analytics	Inmon's early focus was on "departmental analysis", which was experientially true for real-world organizational issues, such as political and funding expediency. Later work expanded concepts to common-analytic needs crossing departmental boundaries.

from The DAMA Guide to the Data Management Body of Knowledge © 2009 by DAMA International

