Building an Agile Metadata Strategy: It’s not as Scary as you Think

Donna Burbank
Global Data Strategy Ltd.

DAMA Rocky Mountain Chapter

October 19th, 2017
Donna Burbank

Donna is a recognised industry expert in information management with over 20 years of experience in data strategy, information management, data modeling, metadata management, and enterprise architecture. Her background is multi-faceted across consulting, product development, product management, brand strategy, marketing, and business leadership.

She is currently the Managing Director at Global Data Strategy, Ltd., an international information management consulting company that specializes in the alignment of business drivers with data-centric technology. In past roles, she has served in key brand strategy and product management roles at CA Technologies and Embarcadero Technologies for several of the leading data management products.

As an active contributor to the data management community, she is a long time DAMA International member, Past President and Advisor to the DAMA Rocky Mountain chapter, and was recently awarded the Excellence in Data Management Award from DAMA International in 2016.

She was on the review committee for the Object Management Group’s Information Management Metamodel (IMM) and the Business Process Modeling Notation (BPMN). Donna is also an analyst at the Boulder BI Train Trust (BBBT) where she provides advices and gains insight on the latest BI and Analytics software in the market.

She has worked with dozens of Fortune 500 companies worldwide in the Americas, Europe, Asia, and Africa and speaks regularly at industry conferences. She has co-authored two books: *Data Modeling for the Business* and *Data Modeling Made Simple with ERwin Data Modeler* and is a regular contributor to industry publications.

She can be reached at donna.burbank@globaldatastrategy.com

Donna is based in Boulder, Colorado, USA. (OK, Eldora, but whatever... 😊)

Follow on Twitter @donnaburbank
Agenda
What we’ll cover today

• Metadata and Agile – Key Definitions & Context
• Building a Metadata Strategy
• An Agile Approach to Metadata Management
• Summary & questions
Metadata is Hotter than ever
A Growing Trend

In a recent DATAVERSITY survey, over 80% of respondents stated that:
Metadata is as important, if not more important, than in the past.
### Metadata is the “Who, What, Where, Why, When & How” of Data

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

• Metadata exists in many sources across & beyond the organization.
What is Agile?

The Agile Manifesto

We are uncovering better ways of developing software by doing it and helping others do it. Through this work we have come to value:

• **Individuals and interactions** over processes and tools
• **Working software** over comprehensive documentation
• **Customer collaboration** over contract negotiation
• **Responding to change** over following a plan

That is, while there is value in the items on the right, we value the items on the left more.

Agile Manifesto courtesy of [http://agilemanifesto.org/](http://agilemanifesto.org/)
Capital “Agile: vs. Lowercase “agile”

• There is the “Agile” design methodology, and then there is just the plain, old meaning of “agile”.

The Agile Manifesto:
• Individuals and interactions over processes and tools
• Working software over comprehensive documentation
• Customer collaboration over contract negotiation
• Responding to change over following a plan

Agile:
Adjective
1. Quick and well-coordinated in movement.
2. Active
3. Marked by an ability to think quickly

courtesy of http://www.dictionary.com/
Agility is not a new concept

1947

2015
The Sages Agree

"The journey of a thousand miles begins with a single step." - Lao Tzu (circa 500 BC)

“Inch by inch, everything’s a cinch. Yard by yard, everything is hard.” - John Bytheway

“A stitch in time saves nine.” - proverb

“If you don’t have time to do it right, do you have time to do it again?” – numerous sources

“Just do it” - Nike

“Implement your metadata project in small, incremental steps, creating ‘quick wins’ that build to a longer-term sustainable architecture.” – Donna Burbank
Find a Balance in Implementing a Metadata Strategy

Focus on Business Value

• Find the Right Balance
  • Data Architecture & Metadata projects can have the reputation for being overly “academic”, long, expensive, etc.
  • No architecture at all can cause chaos.
  • When done correctly, metadata management helps improve efficiency and better align with business priorities

Too Academic, nothing gets done

Business Value

Too “Wild West”, nothing gets done - chaos
Discussion: Where is Your Organization on this Spectrum? “Analysis Paralysis” or “Wild West”?  

• Let’s do a “current state maturity assessment”  
  • Is your organization A, B, C, or D or E? 

Too Academic, nothing gets done 

Too “Wild West”, nothing gets done - chaos 

Business Value
So How Do You Make Sense of It All?

• With the amount of data sources available and stakeholders involved, creating a data strategy can be a daunting task.

• It’s critical to create a Metadata Strategy that:
  • is agile and provides solid results
  • manages the complexity of today’s data ecosystem
  • is sustainable both architecturally & organizationally
Building a Metadata Strategy
Metadata is Part of a Larger Enterprise Landscape
A Successful Data Strategy Requires Many Inter-related Disciplines

<table>
<thead>
<tr>
<th>“Top-Down” alignment with business priorities</th>
<th>Business Strategy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Managing the people, process, policies &amp; culture around data</td>
<td>Data Strategy</td>
</tr>
<tr>
<td>Leveraging &amp; managing data for strategic advantage</td>
<td>Data Governance</td>
</tr>
<tr>
<td>Coordinating &amp; integrating disparate data sources</td>
<td>People</td>
</tr>
<tr>
<td>“Bottom-Up” management &amp; inventory of data sources</td>
<td>Process</td>
</tr>
</tbody>
</table>

Data Governance

- Master Data Management
- Data Warehousing
- Business Intelligence
- Big Data Analytics
- Data Quality
- Data Architecture & Modeling

Data Asset Planning & Inventory

- Databases
- Big Data
- Unstructured Data
- Semi-Structured Data
- Document & Content Mgt.
## Key Components of Metadata Management

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

Business Drivers

**External Drivers**
- Digital Self Service
- Increasing Regulatory Pressures
- Online Community & Social Media
- Community Building

**Internal Drivers**
- Targeted Marketing
- Brand Reputation
- 360 View of Customer
- Efficient IT

Stakeholder Challenges

1. Lack of Business Alignment
   - Data spend not aligned to Business Plans
   - Business users not involved with data

2. Integrating Data
   - Siloed systems
   - No common view of key information

3. Data Quality Issues
   - Bad customer info causing Brand damage
   - Completeness & Accuracy Needed

4. Cost of Data Management
   - Manual entry increases costs
   - System redundancy
   - No reuse or standards

5. No Audit Trails
   - No lineage of changes
   - Fines had been levied in past for lack of compliance

6. Big Data Exploitation
   - Exploiting Unstructured Data
   - Access to External & Social Data

Metadata Capability

**Metadata Strategy**

**Metadata Capture & Storage**

**Metadata Integration & Publication**

**Metadata Management & Governance**

Shows “Heat Map” of Priorities
The “Elevator Pitch”
How Would you Describe Your Project to the CEO in 2 minutes?

I’m working on a project to rationalize metadata across data sources to ensure consistency...

Zzzzzzzz...

I’m working on a project to get a more complete view of customers for the big online marketing campaign...

Interesting!

VS.
Stakeholders: Who Uses Metadata?

- Metadata is used and created by a wide range of roles across the organization.
- Business users are key users.
Metadata is Needed by Business Stakeholders
Making business decisions on accurate and well-understood data

80% of users of metadata are from the business, according to the recent DATAVERSITY survey.

“Metadata helps both IT and business users understand the data they are working with. Without Metadata, the organization is at risk for making decisions based on the wrong data.”
• It’s important to speak with a wide range of roles across the organization.
  • Business & IT
  • Cross-functional teams (Marketing, Finance, Analytics, etc, etc.)
• Understand key opportunities & challenges.
• Recruit allies & volunteers (and identify those you still need to convince. 😊 )
Stakeholder Feedback

• Determine key business issues & drivers through direct feedback.

- There is limited ownership or enforcement of common practices and standards across the projects
- We have 15 customer databases – with many duplications.
- There was an error in reporting products by customer & region that was noticed by upper management.
- Where do I go to get the definition of “default banking standard”?
- I didn’t know we had any documented data standards
- I just joined the company and don’t understand all of the acronyms!
- $12m has been spent on projects to clean up the data over the past 2-3 years
- Key subject matter experts are relied upon to review detailed data from various systems to ensure accuracy.
- I need a central, accurate view of all my customers worldwide.
- What are the data structures used in the application?
Inventory & Usage Mapping

- It’s also important to determine which teams are using these technologies to create a “heat map” of usage & priority.
**Defining an Actionable Roadmap**

**Maximize the Benefit to the Organization**

- Develop a detailed roadmap that is both actionable and realistic
  - Show quick-wins, while building to a longer-term goal
  - Balance Business Priorities with Data Management Maturity
  - Focus on projects that benefit multiple stakeholders

<table>
<thead>
<tr>
<th>Initiatives</th>
<th>H1 '17</th>
<th>H2 '17</th>
<th>H2 '18</th>
<th>H2 '18</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategy Development</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Governance Lineage for Privacy Rules</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Business Glossary Population &amp; Publication</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Warehouse Metadata</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social Media Metadata</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Open Data Publication</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IoT Integration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ongoing</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Integrated Customer View**

- Marketing
- Customer Support
- Sales
- Executive Team
An Agile Approach to Metadata Management
The following are common architectural options for metadata management within & between organizations.

- There is no “one size fits all” approach.
- They can be used together within the same organization.
Automation: Metadata Discovery Tools

- Metadata Discovery Tools extract metadata from source systems, and rationalize them to a common metamodel and storage facility.
- Automations saves manual time & effort – increasing agility
Automation: Data Lineage

• In the data warehouse example below, metadata for CUSTOMER exists in a number tools & data stores.
• This lineage can be automated in most metadata repositories (and some modeling tools).
  • Many can auto-detected metadata based on content (e.g. This looks like a SSN)
  • Custom mapping & matching can also be defined based on business rules.
Data Models can provide “Just Enough” Metadata Management

- While data modeling tools are not metadata repositories, nor designed to be, they offer many features shared with these repository solutions:
  - Metadata storage, Data lineage visualization, Business Glossary, Integration with BI tools, ETL tools, etc.
  - Metadata repositories have a broader range metadata sources & dedicated metadata management support.
  - And Data Modeling tools, of course, have the added benefit of doing data modeling! 😊
    - And the benefit is that much of the needed metadata is in these data models.

<table>
<thead>
<tr>
<th></th>
<th>Metadata Storage</th>
<th>Metadata Lifecycle &amp; Versioning</th>
<th>Data Lineage Visualization</th>
<th>Business Glossary</th>
<th>Data Modeling</th>
<th>Metadata Discovery &amp; Integration w/ Other Tools</th>
<th>Customizable Metamodel</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Modeling Tools</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(e.g. Erwin, SAP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PowerDesigner, Idera</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ER/Studio)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Metadata Repositories</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(e.g. ASG, Adaptive, DAG)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Data Governance Tools</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(e.g. Collibra, Diaku)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Spreadsheets</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data Models are a Good Source of Metadata

- Data Models are another good source of both business & technical metadata for relational databases.
- They store structural metadata as well as business rules & definitions.

**Technical Metadata**

<table>
<thead>
<tr>
<th>Customer</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer_ID</td>
<td>CHAR(18) NOT NULL</td>
</tr>
<tr>
<td>First Name</td>
<td>CHAR(18) NOT NULL</td>
</tr>
<tr>
<td>Last Name</td>
<td>CHAR(18) NOT NULL</td>
</tr>
<tr>
<td>City</td>
<td>CHAR(18) NULL</td>
</tr>
<tr>
<td>Date Purchased</td>
<td>CHAR(18) NULL</td>
</tr>
</tbody>
</table>

**Business Metadata**

A CUSTOMER is a person or organization who has rented a movie within the past year.
Break the Large Modeling Efforts into Manageable Chunks

Instead of creating large models all at once

Break them into smaller “chunks” / sprints

Customer
A Customer is a person or organization with an active account.

Client
A Client is an individual with an active brokerage account.

Account
An Account is a registered savings option at a retail bank that can either be savings or checking. Securities holdings are not considered an account.

Broker
A broker is a registered agent engaged in the sale of securities holdings.

Salesperson
A salesperson is an internal employee trained to sell retail accounts to individuals or organizations.
The Value of Whiteboarding
It’s often helpful to “whiteboard” data models with sticky notes

- Short whiteboard sessions with key stakeholders can flesh out key metadata definitions & scope in a short period of time.
- And it can be fun and interactive.
An Agile Approach to Data Modeling

Rapid Development, Rapid Feedback

- Align with Business Priorities
- Create Subject-Area Focused Working Group
- Source Documentation from Related Efforts

- Scope Business Subject Area(s)
- Define core business entities & relationships
- Draft entity definitions

- Reverse Engineer Physical Models for related systems
- Align with project teams for App & System Delivery
- Iterate refine business model based on differing system rules

Focus on Communication & Iteration

- Publish & Communicate
- Top-Down Business Design
- Bottom-Up Technical Review
- Iterate & Refine

- Fail fast for quick correction and ultimate solid model delivery (Wider Enterprise)
- Utilizing the Agile Sprint approach for constant team and business feedback for quick results (Core Working Group)
A little metadata up-front prevents headaches down the road

- It’s often tempting to skip metadata documentation because it’s “faster”
- But...long-term, it’s ultimately longer as errors and inconsistencies need to be fixed as a result.

“If you don’t have time to do it right, do you have time to do it again?”
Integrating Metadata Management Into the Agile Lifecycle

- Integrating Metadata checkpoints & activities into the Agile development lifecycle helps proactively manage data-related issues before and during development, rather than reactively after the fact.
- Below is a high-level overview of the types of metadata-related questions that can be asked by team members along the various phases of the Agile development lifecycle.

- What are our agreed definitions for core concepts (e.g. Active Account?)
- How will we implement our core data requirements?
- How will we resolve any Privacy concerns?
- Etc.

- Are there any new data requirements for this concept?
- Are key business concepts & terms clearly understood?
- Are there any Privacy concerns?
- Etc.

- Are there common standards that can be reused?
- How do I publish & share my work with others?
- Etc.

- Are there overlaps or conflicts in data usage or design?
- Are other teams defining & using terms differently?
- Etc.

- What are our agreed definitions for core concepts (e.g. Active Account?)
- How will we implement our core data requirements?
- How will we resolve any Privacy concerns?
- Etc.

- Are there any new data requirements for this concept?
- Are key business concepts & terms clearly understood?
- Are there any Privacy concerns?
- Etc.

- Are there common standards that can be reused?
- How do I publish & share my work with others?
- Etc.

- Are there overlaps or conflicts in data usage or design?
- Are other teams defining & using terms differently?
- Etc.

- What are our agreed definitions for core concepts (e.g. Active Account?)
- How will we implement our core data requirements?
- How will we resolve any Privacy concerns?
- Etc.
Metadata Publication & Reporting – Business Glossary

• A Business Glossary is a common way to publish business terms & their definitions.
• When sourced from a common repository, these terms are integrated with the wider data landscape.

<table>
<thead>
<tr>
<th>Business Term</th>
<th>Abbreviation</th>
<th>Definition</th>
<th>Data Steward</th>
<th>Security Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>BFPO Number</td>
<td>BFPO Num</td>
<td>BFPO Number is for British Forces Postal Office. It can be used in UK and overseas addresses.</td>
<td>Accounting</td>
<td>Unclassified</td>
</tr>
<tr>
<td>Interest</td>
<td>Int</td>
<td>The growth in capital of a monetary investment</td>
<td>Finance</td>
<td>Unclassified</td>
</tr>
<tr>
<td>PO Box</td>
<td>POB</td>
<td>A numbered box in a post office assigned to a person or organization, where mail for them is kept until collected</td>
<td>Accounting</td>
<td>Unclassified</td>
</tr>
</tbody>
</table>

A feedback mechanism is important to gather valuable input & updates from users.
Human Metadata
Avoid the dreaded “I just know”

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

• Many metadata & glossary vendors are embracing the concept of “crowdsourcing”.
  • Open editing
  • Popularity & Usage Rankings
  • Dynamically changing
• i.e. The Wikipedia vs. Encyclopedia approach … or Agile Approach

**Encyclopedia**
• Created by a few, then published for read-only.
• Single source of “vetted” truth.
• Static

**Wikipedia**
• Created by a by many, edited by many.
• Eventual consistency with multiple inputs.
• Dynamic
Crowdsourcing is Agile

• Crowdsourcing metadata definitions aligns with the Agile Manifesto

The Agile Manifesto:
• Individuals and interactions over processes and tools
• Working software over comprehensive documentation
• Customer collaboration over contract negotiation
• Responding to change over following a plan

What do you see as the Pros & Cons of this approach?
Summary

• Metadata is more important than ever
• Documenting metadata is both Agile and more agile
• Align metadata with critical business objectives and identify “quick wins”
• Use small “sprints” to create metadata – not all at once
• Have fun! Metadata is for the cool kids.
About Global Data Strategy, Ltd

Data-Driven Business Transformation

• Global Data Strategy is an international information management consulting company that specializes in the alignment of business drivers with data-centric technology.

• Our passion is data, and helping organizations enrich their business opportunities through data and information.

• Our core values center around providing solutions that are:
  • **Business-Driven**: We put the needs of your business first, before we look at any technology solution.
  • **Clear & Relevant**: We provide clear explanations using real-world examples.
  • **Customized & Right-Sized**: Our implementations are based on the unique needs of your organization’s size, corporate culture, and geography.
  • **High Quality & Technically Precise**: We pride ourselves in excellence of execution, with years of technical expertise in the industry.

Visit [www.globaldatastrategy.com](http://www.globaldatastrategy.com) for more information
Global Data Strategy, Ltd

We’re Looking for a few good consultants

• We’re looking to fill the following opportunities:

Data modeling
• Immediate opportunity
• Short term initial project, with opportunity for follow-on
• East Coast Client
• Using the Agile approach

Data Warehousing & BI
• Starting January 2018
• Long term project, 1+ years
• SQL Server stack + erwin
• Midwest client
• Non-profit helping families & children
Contact Info

• Email: donna.burbank@globaldatastrategy.com
• Twitter: @donnaburbank
  @GlobalDataStrat
• Website: www.globaldatastrategy.com
• Company Linkedin: https://www.linkedin.com/company/global-data-strategy-ltd
• Personal Linkedin: https://www.linkedin.com/in/donnaburbank
White Paper: Emerging Trends in Metadata Management

Free Download

• Available https://globaldatastrategy.com/
  • Under Resources - Whitepapers
DATAVERSITY Training Center
Online Training Courses

Metadata Management Course

• Learn the basics of Metadata Management and practical tips on how to apply metadata management in the real world. This online course hosted by DATAVERSITY provides a series of six courses including:
  • What is Metadata
  • The Business Value of Metadata
  • Sources of Metadata
  • Metamodels and Metadata Standards
  • Metadata Architecture, Integration, and Storage
  • Metadata Strategy and Implementation

• Purchase all six courses for $399 or individually at $79 each. Register here

• Use code “GDS” for 20% off

• Other courses available on Data Governance, Data Quality, etc.

Visit: http://training.dataversity.net/lms/
Questions?
Thoughts? Ideas?