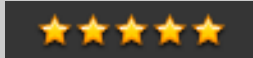


Redefining Metadata

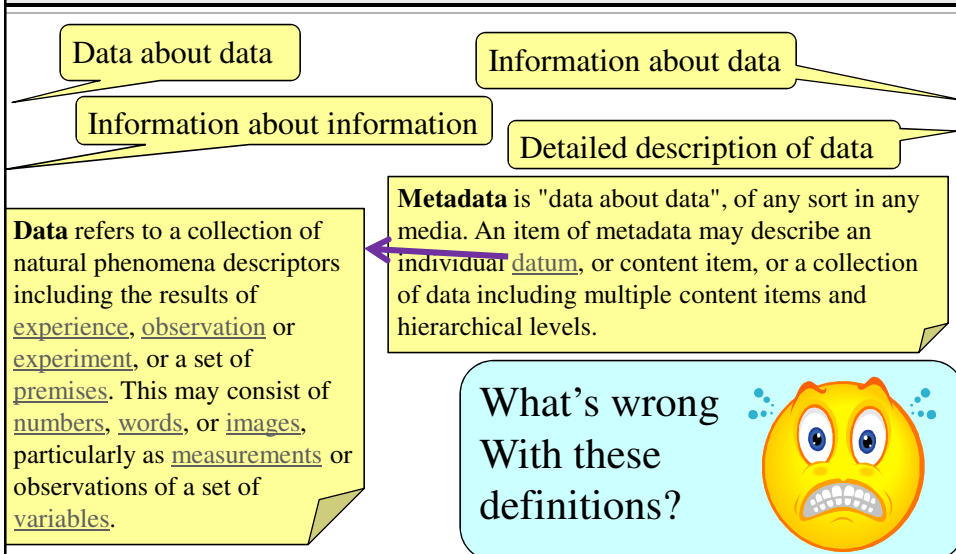
Steve Hoberman
 me@stevehoberman.com
 www.stevhoberman.com

Data or Metadata?		
	Data	Metadata
42	<input type="checkbox"/>	<input type="checkbox"/>
“Begin with the end in mind”	<input type="checkbox"/>	<input type="checkbox"/>
<i>Data Modeling Made Simple</i>	<input type="checkbox"/>	<input type="checkbox"/>
Vendor Name	<input type="checkbox"/>	<input type="checkbox"/>
	<input type="checkbox"/>	<input type="checkbox"/>

You will learn




Our accepted definitions of metadata



My thoughts on metadata

Metadata is text that describes what the audience needs to see.



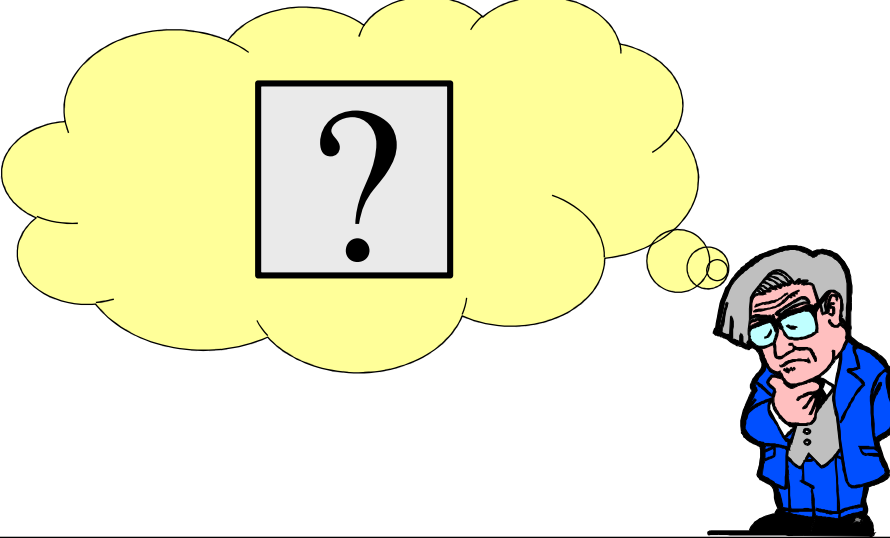
What is metadata?

Metadata is text that describes what the audience needs to see. A particular context can turn what we traditionally consider data into metadata. Metadata is important because it aids in clarifying and finding the actual data. There are four types of metadata: business, storage, process, and project.

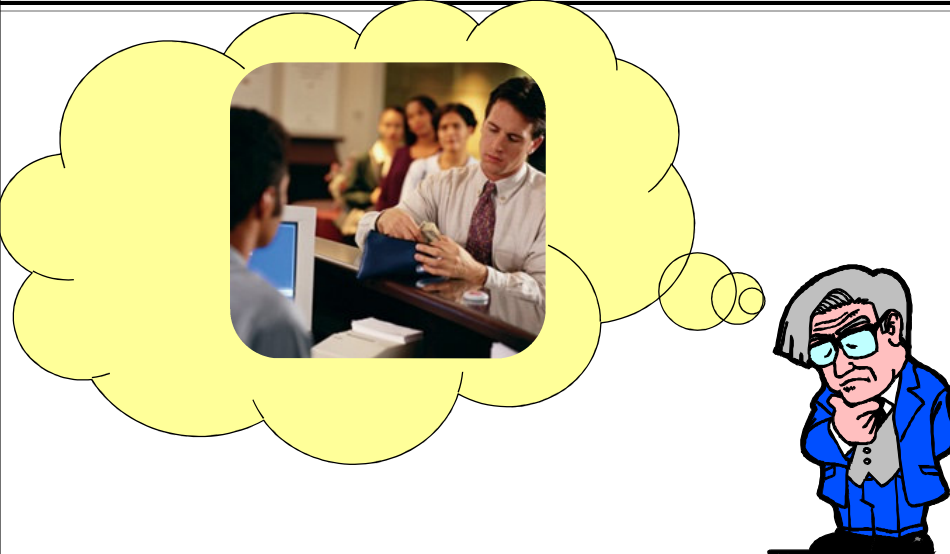
Examples:

- Business – Definitions, tags, calculations, business names
- Storage – Database names, formats, nullability, volumetrics
- Process – Source/target mappings, transformation logic
- Project – Functional Requirements, Zachman Framework

A context game
Write one sentence defining the geographic shape 'Square'



A context game
Write one sentence defining 'Customer'



Context can turn data into metadata

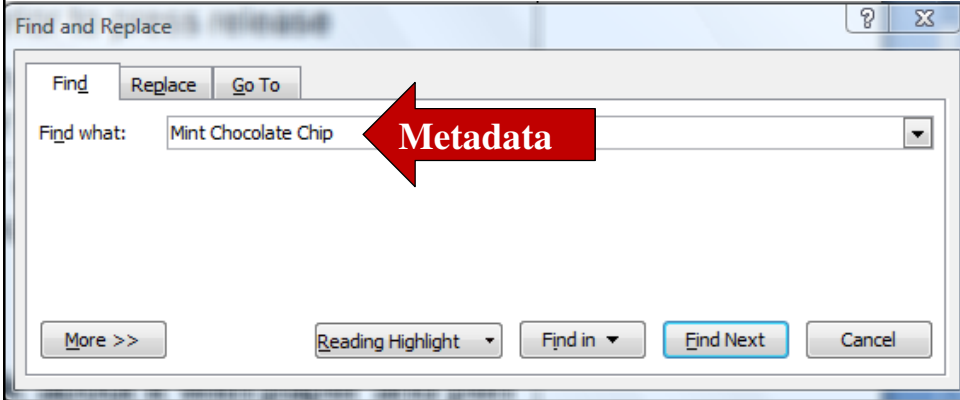
There used to be a difference between data and metadata. Data was the suitcase and metadata was the name tag on it...Data was the contents of the book and metadata was the Dewey Decimal number on its spine...Now all is metadata. Data is all surface and no insides. It's all handles and no suitcase. It's a folder whose content is just another label. It's all sticker and no bumper.
David Weinberger

This is no longer the Age of Information. It's the Age of Metadata.
Blogger

Context can turn data into metadata

I really loved all of their flavors, but by far my favorite is Mint Chocolate Chip.

Data ←



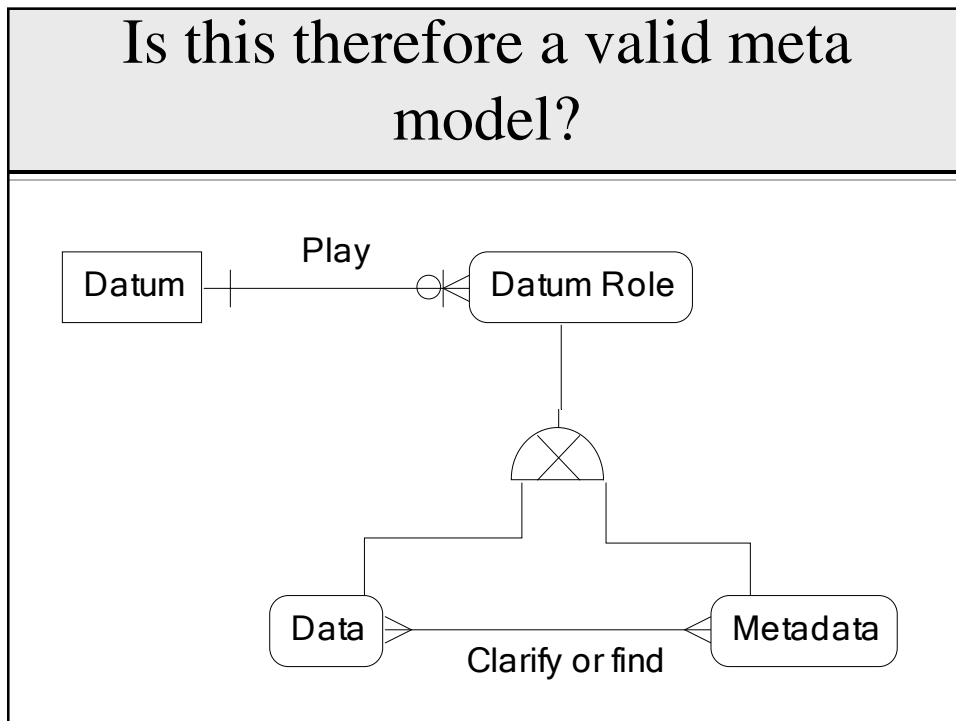
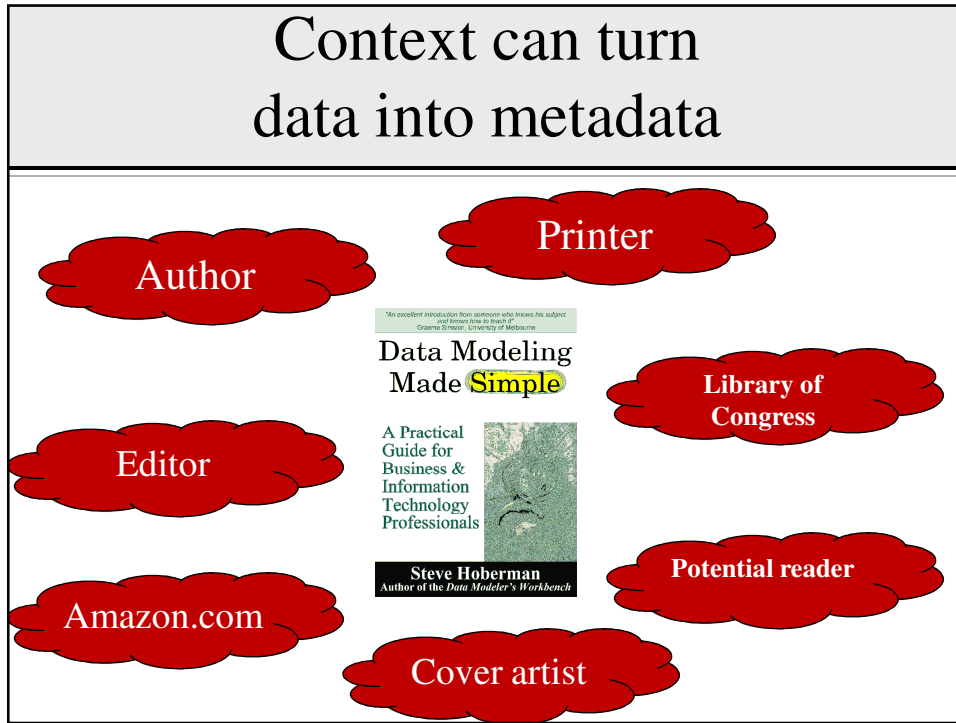
Find and Replace

Find Replace Go To

Find what: Mint Chocolate Chip


Metadata ←

More >> Reading Highlight Find in Find Next Cancel



Do you agree with this statement from Wikipedia?

As for most people the difference between data and information is merely a philosophical one of no relevance in practical use.



Why is it important to know when data is playing the role of metadata?

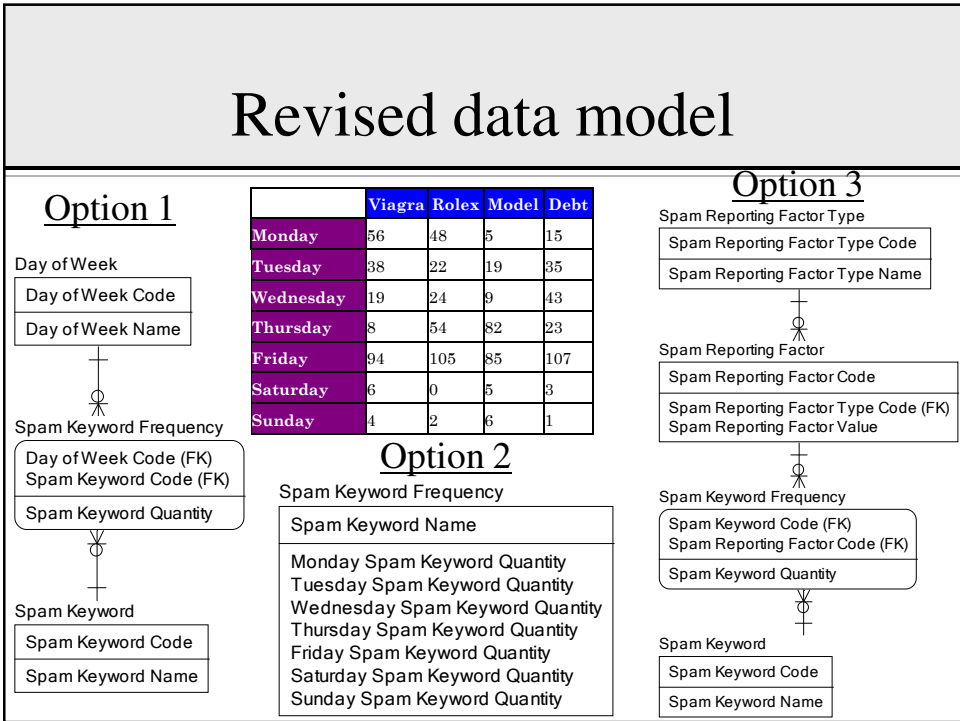
- Analysts ask the right questions and capture the right requirements
- Data quality taken more seriously
- Well-defined business processes
- Example
 - I need to see email subject text

Here you go!

Email
Email identifier
Email From Address
Email To Address
Email Sent Date
Email Subject Line Text

Thank you data modeler, but how can I produce this report?

	Viagra	Rolex	Model	Debt
Monday	56	48	5	15
Tuesday	38	22	19	35
Wednesday	19	24	9	43
Thursday	8	54	82	23
Friday	94	105	85	107
Saturday	6	0	5	3
Sunday	4	2	6	1



Top 5 challenges with data becoming metadata

1. Greater reliance
2. Greater responsibilities
3. New roles
4. Additional data quality checks
5. Less rewards



Greater reliance

- More searching
- “Googling” becomes a common term in requirements documents
- Less patience to finding information
- “Document management meets dimensionality”
- Electronic discovery



Greater reliance

```

graph LR
    C[Consumer] ---|Place| Co[Complaint]
    Co ---|"Be the subject of"| P[Product]
    
```

I want to see what each consumer has to say on our products.

Greater responsibilities

- Additional requirements that need to be mapped and modeled
 - Tagging
 - Ontologies and taxonomies
 - Privacy rules
- Influencing workflow
- More development effort as well
- One of many examples: Imagine the effort to resolve new integration issues
 - “Does everyone define this hierarchy the same way?”

Greater responsibilities



These are my daughters.

No, these are my granddaughters.

No, you are both wrong! These are sisters.

Anyway, this is a garage sale.

We call it a boot sale.

Actually, it's a yard sale.

New roles

Which of the following are currently roles within your organization?

- A. Metadata Steward
- B. Ontologist
- C. Cataloger
- D. Tagger
- E. None of the above



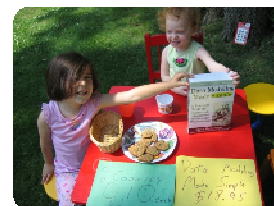
New roles

- How many of you have budgets set up for metadata management?
- How many of you have lots of free time at work?
- If we are to play these roles, need
 - Different mindset
 - Greater exposure to industry standards
 - More time!



Additional data quality checks

- More than just having the correct data – also need to ensure the correct data / metadata relationship
 - Is the tagging of this photo correct?
 - Is 'Garage Sale' correct?
 - Does 'Garage Sale' belong to event or location?
- Need data creation and change processes



Less rewards

- Lack of recognition for a job well-done
- Those in metadata management are traditionally the unsung heroes
- There will be even more opportunities to become unsung heroes

Congratulations to successfully achieving another milestone on our process and system transformation journey. The impact of any Data project gets largely ignored, especially if it almost becomes a 'non-event' to the business. Congratulations again to you and the entire team.



What we can do now

- Need to treat data like metadata
 - Formal change process
 - Responsibility
 - Business/IT partnership
 - Industry standards
- Study how search engines like Google work
 - infolab.stanford.edu/~backrub/google.html
- Introduce meaningful metadata metrics
 - % Tagging Complete
 - % Tagging Signed Off
 - % Tagging Integrated

(but don't use the word 'metadata')
- Create Metadata Centers of Excellence



Missing the requirement

- Partner gives employee task of organizing his documents
- After 3 weeks, partner is shown shelf containing 11 binders
- Content is there with no way to access
 - “I want to see all the documents containing the term ‘treaty’.”
 - “I want to see all the documents for Company ABC.”
 - “I want to see all the documents from April 2008”
 - “I want to see all of the documents of type ‘Tax Deduction’.”



Build a SAM for this document management system

Then...

- For each concept on your SAM, identify the role that is responsible for this data

Concept	Role

Key points

- Metadata is text that describes what the audience needs to see
- Context can turn data into metadata
- Top 5 challenges
 - Greater reliance
 - Greater responsibilities
 - New roles
 - Additional data quality checks
 - Less rewards
- What we can do now
 - Treat data like metadata
 - Study search engines
 - Introduce metadata metrics
 - Create Metadata Centers of Excellence

