Metadata & the Metadata Repository
Michael McCarthy

Content

• What is Metadata?
• The Oxford English Dictionary Analogy
• Primary users of Metadata
  – Technical & Operations
  – Business
• What Metadata is Required?
  • Categories of Metadata
    – Ownership
    – Origin, Derivation, Physical Characteristics
    – Quality
    – Security
• What is a Metadata Repository?
• Objectives of a Metadata Repository
• CRM and the Metadata Repository
• Metadata Repository Vendors

What is Metadata?

• Literally, “data about data”.
• Information about any business object or piece of information collected or used by an organisation, such as:
  – The currency of the data
  – The origin of the data
  – Measures of data quality
  – Physical characteristics of the information
  – Owner of the data
• Must be maintained
  – Metadata changes over time.
The Oxford English Dictionary

- The Oxford English Dictionary (OED) is not just a standard dictionary.
- It includes over 500,000 words with their:
  - Meaning
  - History
  - Pronunciation
- It traces the usage of words through a wide range of sources
  - From classical literature and specialist periodicals
  - To film scripts and cookery books

To gain an understanding of metadata and its associated functions we can draw an analogy between the OED and Metadata.

The components involved in the OED are:
- Other Languages - The languages from which English has originated
- Grammar - The rules which form the basis of correct English usage
- Characteristics - Features that make up various instantiations of the English language (eg. British English, American English)
- Meanings - The meanings or definitions behind each word

We can compare these with Metadata components:
- Source System - The source systems from which data is derived
- ETL - The process in which data is extracted, loaded and transformed
- Physical Database - The actual place the data resides
- Profiling - The process of capturing characteristics of data
- Logical Model - The allowable data structure and relationships
- Business Definitions - The business meaning of the data

Metadata Management

With thanks to Bob Mason, ANZ
**Metadata and The Oxford English Dictionary: A Comparison (A)**

![Metadata Diagram](image)

**Metadata and The Oxford English Dictionary: A Comparison (B)**

<table>
<thead>
<tr>
<th>No.</th>
<th>Metadata Function</th>
<th>OED Equivalent Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Source Systems</td>
<td>Other Languages</td>
</tr>
<tr>
<td>2</td>
<td>ETL</td>
<td>History of the Language (the changes)</td>
</tr>
<tr>
<td>3</td>
<td>Physical Database</td>
<td>Spoken English</td>
</tr>
<tr>
<td>4</td>
<td>Data Profiling</td>
<td>Characteristics</td>
</tr>
<tr>
<td>5</td>
<td>Data Modelling</td>
<td>Grammar</td>
</tr>
<tr>
<td>6</td>
<td>Data Definitions</td>
<td>Meaning (Definition)</td>
</tr>
<tr>
<td>7</td>
<td>Metadata Repository</td>
<td>Oxford English Dictionary</td>
</tr>
</tbody>
</table>

**Users of Metadata - Technical**

- Systems developers, e.g.
  - Impact Analysis, e.g.
    - add a new valid value, “X” to Gender
    - Add a new column to Customer entity
Users of Metadata - Operations

- Operations - those personnel responsible for ensuring efficient operation of the system, e.g.
  - Job run statistics
  - Job start and end time
  - Elapsed time
  - Number of records loaded/in error

Users of Metadata - Business

- Business users - the end users of information within the system, e.g.
  - Data quality
  - Data currency
  - Data owner
  - Data Lineage

What Metadata is Required?

- The metadata that is required should meet the following criteria:
  - Makes good business sense
  - Is put to appropriate business use
  - Is accurate
  - Is easy to maintain, i.e.
    - Kept up to date and accurate
    - Discarded when appropriate
  - Is accessible
  - Is cost-effective [to capture and store]
Categories of Metadata - Business Definition

- Business meaning
  - E.g. "C_CUST_LST_TXN_AMT" really means "Customer’s Last Transaction Amount"
- Valid values
  - E.g. For Gender “M” for Male and “F” for Female. Any other value is invalid
- How the business uses and collects the information

Categories of Metadata - Ownership

- Person or Business Unit responsible for maintaining the:
  - Quality of the data
  - Business definition
  - Types of transactions that generate data
  - Processes and procedures for capturing data accurately

Categories of Metadata – Origin, Derivation & Physical Characteristics

- Source system/s
  - Physical location, e.g. DB2, IMS, SQL Server, Access, Excel
- Input fields
  - Format, e.g. Date, Text, Character, Numeric
- Calculations
  - E.g. Inclusive or Exclusive of GST
- Refresh period/frequency
  - E.g. Daily, Weekly, Monthly
- Relationship to other data structures and processes
  - E.g. Gender – “M” and Title “Mr”
- Transformations of the data
  - E.g. If Country Code = “AU”, then Residency = “Local”
Categories of Metadata - Quality

• Complete  
  – E.g. Every column in Gender has a value, no nulls.
• Valid  
  – Every value is “F” or “M”
• Correct  
  – Every “F” is representative of a female customer and every “M” is representative of a male customer.
  – Title, Salutation and Gender are aligned

Categories of Metadata - Security

• Information about constraints on ability to:  
  – Create  
  – Update  
  – Delete  
  – Read, and  
  – Who

What is a Metadata Repository (A)

• An automated store that holds Metadata in a format that is accessible and maintainable and is linked to the data it is describing.
• Can be a virtual repository supported by one or more physical sources (tools interchange).
• Contains descriptions of all tables and columns in the data warehouse and all business rules.
**What is a Metadata Repository (B)?**

- Can be used to generate database schema changes (active repository).
- Supports versioning.
- Describes data lineage.
- Supports different environments:
  - Development
  - Test
  - Production
- Supports reverse engineering

**Objectives of a Metadata Repository**

- Empower business users
- Actively use a shared and integrated (unified) layer of business definitions and technical specifications
- Improve development and support productivity through enhanced data documentation
- Provide a history of object definitions
- Create a sustainable information environment
- Improve the productivity of the design team
- Facilitate improvements in data quality measurement and data conversion efficiency
CRM and the Metadata Repository

- Organisations want to:
  - Be more customer-focused
  - Identify where customers are on their lifetime journey
  - Do more cross-sell and up-sell
  - Do more retention and less acquisition
  - Develop relationships with their profitable customers
- All require data and information about the data, i.e. Metadata.

Metadata Repository Vendors

- Data Advantage Group - www.dag.com/index.asp
- MetaMatrix - www.metamatrix.com
- ASG Worldwide Headquarters - www.asg.com
- Ascential Software - www.ascential.com (see IBM)
- Computer Associates - www.ca.com
- SAS - www.sas.com
- Business Objects - www.businessobjects.com
- Oracle - otn.oracle.com
- Informatica - www.informatica.com
- IBM – www.ibm.com

Conclusion

- Questions
- Thank you