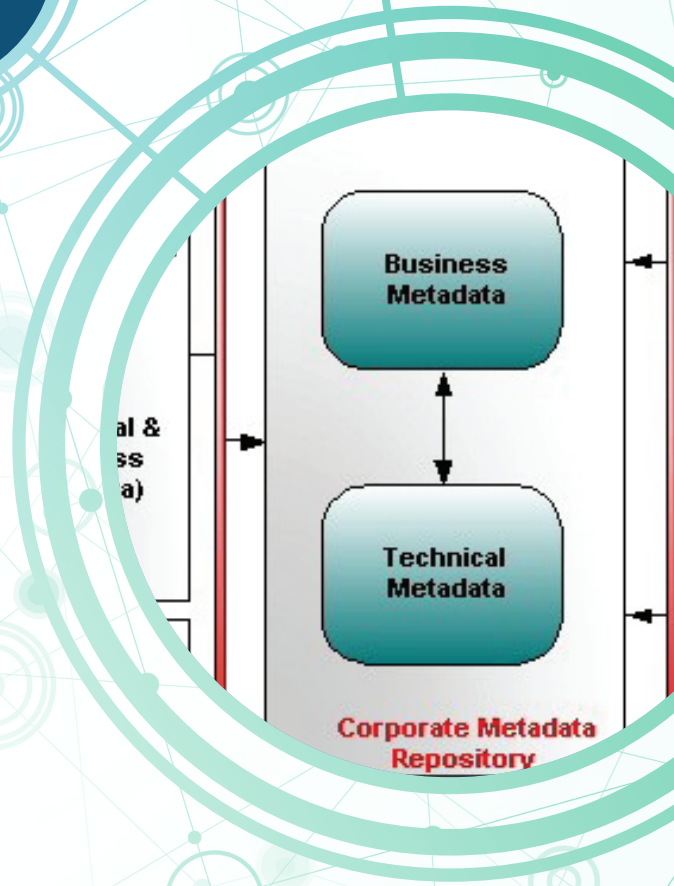


Unissant™

Unissant White Paper Series



Corporate Meta-Data Building Blocks—An Introduction



What is Meta-data?

Meta-data in most simple terms means “structured data about data” or “data that describes something (that may or may not itself be data)”.

The prefix “meta” comes from the Greek and can indicate change, as in metamorphosis; or it can mean beyond or after, as in metaphysics. In information technology usage, the word metadata has come to be used as a definition or description of data: a small indicator that encompasses and points to a larger piece of information. The library card catalog is the standard metaphor for metadata: each card represented and led the user to a much larger body of information, the book or other item cataloged.

Data is a valuable corporate asset, which outlasts applications and processes. Analysis of data allows organizations to learn and grow. Metadata has been around since the first program was written. The word metadata was first recorded in the dictionary around 1980’s but its usage goes back all the way to 1960’s.

Why do we need to describe things?

- “discover”: to find out that things exist
- “locate”: to find out where things are
- “request”: to ask for something
- “access”: to get something – so we can use it

Metadata provides us with a reference to useful data, and provides us with definitions that allow us to understand that data: its definition, its usages, its ownership, and its quality. Metadata describes the processes and tools that maintain and act on that data. So metadata goes beyond the traditional data dictionary or card catalog, and can describe the entire environment, which we have built to collect, maintain, and deploy our data assets.

Why do we need metadata?

- Poor documentation often leads to loss of critical information
- To help you publicize and support the data or your organization
- So you won’t forget how you collected and processed your own data
- So information is not lost when an employee leaves
- So the data can be used again in the future
- So you can tell if you need other’s data and how to use it
- To help allow your organization share data in a consistent way without duplication
- To preserve the value of the data saving it from being useless due to changes to the data over time
- To help new employees understand the organizations data with a smaller than normal learning curve

Some sources for Metadata in an organization:

- Conceptual, Logical and Physical Data Models
- Business Definitions
- Business Processes
- Technical Definitions
- Business Rules

- Transformation Rules
- Attribute Characteristics – field type, field size, etc.
- Data Source System(s)
- Data Creator
- Data Purpose
- Data Quality Rules
- Data Quality Issues and more

In conclusion, metadata has broad applicability across the enterprise. There is virtually no process in the entire IT organization that can't benefit from metadata. In the next article I will try to dive deep into the types and sources of metadata.

Business Meta-Data Building Blocks

There are two broad categories of meta-data: technical and business. Technical meta-data is the description of the data needed by various tools to store, manipulate, or move data. These tools include relational databases, application development tools, database query tools, data modeling tools, data extraction tools, online analytical processing (OLAP) tools, and data mining tools.

Business meta-data is the description of the data needed by business users to understand the business context and meaning of the data. Though acceptance of meta-data has been slow by the business community, it is a fact that business meta-data is equally important as technical meta-data. In simple terms business meta-data allows business users to use the data more effectively. The question is what should the starting point to start collecting business meta-data be and how should it be integrated with technical meta-data. In this multi part article we will try and focus on various areas of business meta-data and how it integrates with technical meta-data.

Business meta-data can include but is not limited to information about corporate business elements, facts, processes, sub-processes or activities, data quality metrics, business rules, legal aspects of business elements, business events, associations of the business elements to various roles, unstructured information (captured in e-mails, word documents etc.) and much more. As a starting point let's discuss the very basic kind of business meta-data surrounding business element definitions and facts and how it integrates with the technical meta-data.

So what is meant by business elements? Business elements are key business attributes as identified by the business processes and compiled in various subject areas of logical data models. These elements are compiled from extensive experience of various business architects/analysts.

So what is a good source of these elements? In large organizations there are multiple logical data models and business process models, business elements are a collection of all the unique logical data elements pulled out of these models. Ideally, these elements should have consistent definitions and characteristics and should be defined once at an enterprise level. Some organizations maintain business information dictionaries or business definition dictionaries to document these. The definitions can also be extracted from industry standards wherever available. For example in the secondary mortgage business Mortgage Industry Standards Maintenance Organization (MISMO), whose mission is to develop, promote, and maintain voluntary electronic commerce standards for the mortgage industry.

For a business element some additional characteristics should be captured:

Business Element Name(s) short & long

Well-defined short and long business names

Business Definition

Business definition of an element as used by the organization. This definition could be extracted from an industry standard (if available). In cases where the industry standard definition is not available, the definition should be carefully drafted. There are several good articles that have documented guidelines on how to write business element definitions. In situations where the internal definition is not same as the industry definition maybe because the external definition had changed or was introduced at a later stage, both these definitions should be made available in the central repository with the appropriate source identified.

Element Alias

Business element aliases as used by different applications or systems. These aliases could be Element Screen Labels or XML Tags or Reporting Labels etc.

Standard Acronym

Most commonly used acronym (if any) for the business element

Element Synonyms

Business element synonyms as used by different divisions or departments.

Element Criticality

Certain business elements are critical to the organization in terms of financial reporting or key performance indicators (KPI's). These elements should be flagged critical for many reasons including addressing compliance issues. There are many more characteristics that can be tracked in terms of criticality e.g. source of critical element, reason for criticality, effective date, expiration date etc.

Frequency of use

How often is the element reported on or used in various associated processes.

Confidence level

Business element attribute used to measure the organization's faith in a specific data field. Usually expressed in a three level scale (poor, moderate, high).

Business Source

Which business area or division identified the need for this business element.

Security Characteristics

Is the element classified as "secure". Furthermore is the classification internal or external to the organization. This classification helps determine what level of security is required for the data element and any transactions pertaining to it. Example of the values can be: 1 = Highly Confidential (should be restricted for both internal and external viewing); 2 = Moderately Confidential (should be highly restricted for external parties and not as restricted for internal parties); 3 = Public (anyone can view or read the data).

Promoted Physical Name & Characteristics

Promoted physical name and characteristics that should be used by the data architects to ensure all physical models are consistent.

Element Creation Date

Date when element was first introduced in the organizations vocabulary of business elements.

Element Categorization

Element categorization with buckets defined as “Technical Attribute”, “Risk Management Attribute”, “Operational Attribute”, “Legal Attribute” etc.

Valid Values

A common problem in large organizations is inconsistent set of valid values relating to enumerated elements. Several values and codes are defined across departments or divisions, which cause confusion. E.g. List of Product(s) or product codes should be defined in the central repository once. Any new data model should extract these values from the central repository. Sometimes it is difficult to define a common set of values, as there might be a requirement that the same code might need to be used in different departments or divisions with different characteristics. In these cases a source name (e.g. department or division) should be tracked in the central repository.

Element Issues

There can be multiple potential issues associated with an element at any given time. E.g. the definition of a particular has been changed in the industry and the company still is using the old definition or a particular required data element has a high percentage of “nulls” in a database, which was identified as the system of record for that element.

These elements should be flagged to have issues alerting users to be careful while using these elements.

Element Risks

Element level risks can be tracked and can be categorized as high, low or medium or level 1, level 2, level 3 depending on the industry. The meta-meta data about the categories should also be tracked in the CMDR e.g. if we say a particular element is high risk what does it really mean to a technical or business user.

Associated roles

Roles including, element owner(s), advocate, steward, subject mater expert(s) (SME) should be tracked at an element level. These roles are critical for maintaining the accuracy of what these elements mean and how they are used in the organization.

An owner is one who actually creates the data pertaining to the elements he is responsible for and has the right to change the data.

A steward is one who is called upon to exercise responsible care over possessions entrusted to him or her (as per Webster’s dictionary). He or she does not own these possessions. Stewards are responsible of ensuring the definitions are complete and accurate. Ideally there should be only one steward per element. Stewards can also be identified at subject area levels in which

case they own all the business elements in that subject area.

A SME or subject matter expert has the most knowledge as to how the elements are being or should be used in various applications.

Element Physical Implementations

In the ideal world, all physical data models should have been generated from logical data models, and all logical models should be in sync in terms of logical names and characteristics. However, in the real world, same business elements are implemented in multiple logical data models with different names and characteristics. Similarly various physical data models define same business elements with different physical names and physical characteristics. Some data architects take the short cut and skip the logical models altogether. All these scenarios mandate a need to associate the business elements to all its physical instances and implementations in the central repository. This can be a challenging task but is also a key integration point between business meta-data and technical meta-data. Assuming that all the physical data models were imported in the central repository, associations need to be made in terms of business elements and all its possible physical implementations across all physical data models (or databases, if you load database catalogs in the repository). Some experts refer this as “stitching” and argue that this is an easy step. Let me tell you folks, depending on how bad the situation is on your organization, this can be the most daunting task. It is essential that these associations are accurate to provide proper impact analysis reports from the repository.

**The above-mentioned list is not a complete list but does cover some important aspects of what needs to be tracked for business elements.

At a high level this is where we are at this time:

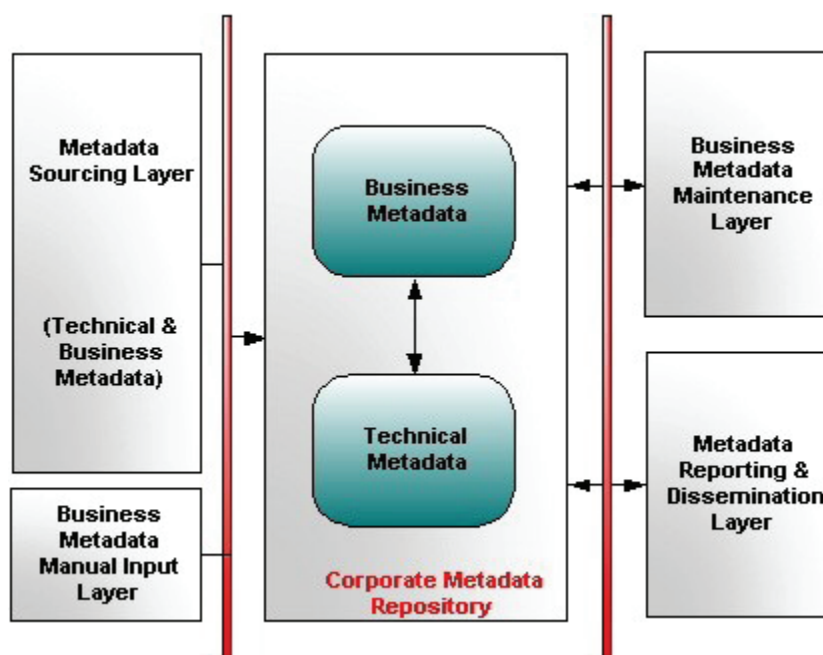


Figure-1: Logical Design Central Meta-data Repository

As depicted in the figure above, both technical and business meta-data are sources from the sourcing layer in our repository. Though business meta-data might exist in various forms, Microsoft Excel spreadsheets, Microsoft Access databases (maintained by shadow IT organizations)

etc. It might also be required to be inputted into the repository directly using a graphical user interface (GUI). The CMDR integrates both types of meta-data and then the client layer is used for reporting and meta-data dissemination. Again a maintenance interface might need to be provided for maintaining the business meta-data. This interface should be strictly for business meta-data, as changes to technical meta-data should be captured in an automated way. Yes an obvious concern would be who has what privilege to change or maintain business meta-data since it is critical to the organization. Let's just say there is a role based security layer that gives access to the interface. This layer could be integrated with your existing corporate security architecture (LDAP).

Integrating business meta-data in the central repository comes with its own set of problems. It is relatively easy to scan, load and maintain technical meta-data, but keeping business meta-data up-to-date and synchronized can be a big challenge, especially if manual interfaces are being used to input and maintain it. Make sure you have a “buy-in” from senior management (both IT and Business) to keep the business meta-data accurate and complete. Ensure a “Change Management” process is in place and followed to document all changes made to the business meta-data.

Remember, a half-baked meta-data solution can quickly turn some users away. When they stop using meta-data, they stop populating the meta-data repository. An incomplete meta-data solution would drive away more users and would result in a vicious circle. Fewer meta-data users result in obsolete meta-data, which erodes its accuracy and completeness, which, in turn, drives more users away.

Bibliography

Milstead, Jessica and Susan Feldman - “Metadata: Cataloging by any other name ...” in Online, Jan/Feb 1999, p24–31. A very lucid description of what metadata is and does, what the term means, and what the challenges are in implementing metadata schemes.

D Gleason – “An Evolutionary Approach to Metadata Management”

About Unissant

Unissant is an advanced data analytics and business transformation services provider with expertise in healthcare and health IT, finance, national security, and energy. The company delivers innovative solutions to assist government agencies and private sector businesses in tackling their biggest challenges. Founded in 2006, Unissant is a prime contractor on various government vehicles such as CIO-SP3, GSA PSS, GSA HealthIT SIN, and GSA 8(a) STARS II and is a CMMI Level 3, ISO 9001 & 27001 certified company headquartered in Herndon, Virginia with a satellite office in San Antonio, Texas. In March 2017, Unissant received the Government Project of the Year award by Small and Emerging Contractors Advisory Forum (SECAF). Unissant has experience with implementing machine learning solutions and full stack implementation on cloud. Unissant is a Select Consulting Partner Public Sector and is leveraging this relationship with AWS to strengthen our ability to support our customers in their Digital Transformation journey. We are already supporting several of our customers in the public sector with their Machine Learning implementations, cloud adoption strategy and implementation.

Copyright

Copyright © 2020 Unissant, Inc. All Rights Reserved.

Restricted Rights Legend

This document may not, in whole or in part, be copied photocopied, reproduced, translated, or reduced to any electronic medium or machine-readable form without prior consent, in writing, from Unissant, Inc.

Information in this document is subject to change without notice and does not represent a commitment on the part of Unissant. This document is provided “as is” without warranty of any kind including without limitation, any warranty of merchantability or fitness for a particular purpose. Further, Unissant does not warrant, guarantee, or make any representations regarding the use, or the results of the use, of the written material in terms of correctness, accuracy, reliability, or otherwise.

All other brand and product names are trademarks of their respective companies.

For additional information on Unissant full range of services, please visit our website at www.unissant.com or call us at 703.889.8500.