

metadata management

ASSIGNMENT TWO



October 16, 2020

jAMIE LU

W0441213

Table of Contents

[Introduction: 2](#_Toc53969959)

[TASK ONE: 3](#_Toc53969960)

[METADATA STRATEGY 3](#_Toc53969961)

[Goals 3](#_Toc53969962)

[Metadata 3](#_Toc53969963)

[Metadata Roles and Responsibilities 4](#_Toc53969964)

[Use of Metadata in the organization 4](#_Toc53969965)

[Metadata Capture 4](#_Toc53969966)

[Quality of metadata sources 4](#_Toc53969967)

[Metadata storage 5](#_Toc53969968)

[Evaluation of metadata products and their capabilities 5](#_Toc53969969)

[Metadata Training 5](#_Toc53969970)

[Priorities 6](#_Toc53969971)

[Business Drivers 6](#_Toc53969972)

[Sources of data 6](#_Toc53969973)

[Metadata Performance 6](#_Toc53969974)

[Technology 7](#_Toc53969975)

[Technical architecture for managed metadata environment (MME) 7](#_Toc53969976)

[Use and effectiveness of the metadata 8](#_Toc53969977)

[TASK TWO: 9](#_Toc53969978)

[DIAGRAM OF METADATA STRATEGY 9](#_Toc53969979)

[TASK THREE: 10](#_Toc53969980)

[ORGANIZATIONAL CHART 10](#_Toc53969981)

[References 11](#_Toc53969982)

# Introduction:

Metadata is everywhere and it is very essential because it is the data about data. It helps provide definition of the data and what the data's context is. Having a metadata strategy for a business or organization is beneficial because it will help in the management of their data. This means that there is more comprehension with regards to the data within the different areas of the business or organization. This will also ensure that there is a map to help each user find, analyze, and retrieve data conveniently.

# TASK ONE:

## METADATA STRATEGY

The business will track metadata to have cohesion in the data in the business. It will add more value into the data. It will be easier to understand it from different perspectives of the business. A metadata strategy will be good to have as well as this will map things out making it easier for users to retrieve, analyze and use data in a meaningful way.

### Goals

* Have ease of access to metadata by creating an easy to comprehend map
* Be able to understand the business drivers and use these to improve the metadata strategy
* Define the roles and responsibilities
* Create a map that is scalable, making growing data easier to store and still able to meet requirements
* Maintain high quality metadata and high level of security
* Have a degree of the value of information
* Recognize the value of data, its metadata, and its components
* Metadata should be able to show what is being stored, its location, and how it is being stored
* Metadata should be able to show transformations has it been through
* Metadata should show that the metadata has been organized to easily retrieve data
* Monitor metadata quality and metrics

### Metadata

Who makes the decisions?

To follow best practice, there will be a **Metadata Administration Team**. This team will comprise of the Data Custodians (the people that will handle data administration (the Data Analyst and Repository Administrator), and the database administrators), the application development programmers, and the Data Stewards. The Metadata administration team would also need to determine what should be metadated, the role of the metadata in the organization and how it will help further goals and determine has access to metadata and to what extent.

The Data stewards will make sure that there is consistency and integrity of data. The Data custodians will make sure that the data has syntactical value. The whole team will organize and discuss while the Metadata stewards will have the final say. This team will establish what the standards and procedures are, define proper use of data and metadata, and will ensure quality control.

Standardization

The standards of Metadata should be determined and established by the Metadata Administration team. They will specify the naming standards, specifically the abbreviations used in the business and systems, the class words, code values, and the business names and definitions. They will also establish what the procedures are for creating or modifying elements and entities. They will also establish how metadata and data is stored.

In creating the standards, they should determine how they will structure the metadata. They can use a standard schema or create one that is shaped based on the needs of the business.

### Metadata Roles and Responsibilities

The **executive sponsor**, this will be the one that understands and communicates what the significance of metadata management is within the business.

The **steering group,** they will be responsible for prioritizing and setting directions for key activities.

The **data stewards,** they are responsible for setting the business definitions and the rules for the key elements of data.

The **Metadata Repository Administrator,** responsible for the management of the administration, the population, and the interfaces of the metadata repository.

The **Metadata Publicist,** responsible for the establishment of reports and publication methods going to the end users.

The **Metadata Consumers**, they are the ones who actively uses the metadata as part of their daily work and are held accountable for using published standards. Examples of Metadata consumers are the data modelers, developers, business user, report developers, etc.

### Use of Metadata in the organization

The usefulness of metadata in the organization must be determined. This will be done through determining where the metadata will be (and will not be) needed or used, and when the metadata is too much or is limited.

### Metadata Capture

To achieve good metadata management, the internal and external sources must be determined. The business must determine what they are specifically trying to capture. This can be done using a combination of data modeling, metadata repositories and data governance tools. These will help the users in evaluation and specifying the Metadata that is captured.[[1]](#footnote-1)

Metadata can also be captured through documented management software. This is a powerful and automated way to do it. While new data is added, the user putting in the data is prompted to classify it using a premade metadata field. This ensures that the necessary amount of information is put in for the data’s metadata (not too much and not limited as well). There will also be a way to optionally add additional metadata. Once the metadata is put in, it will be organized into individual or multiple groups so that users will only have access to the metadata that relates to what area or department they belong to.

### Quality of metadata sources

The quality of metadata sources affects the performance of Metadata. It will affect metadata’s accuracy, completeness, and its consistency. To ensure the quality of metadata sources, the origin of the data, as well as its age must be considered. Any term or tag changes must be considered as well as its relation to the other metadata. It should hold the historical views of the metadata as it evolves through time.

### Metadata storage

The main job of storage is to be able to hold the information and be able to retrieve it when you need it. Both need to be done securely, reliably, and conveniently. Good storage should be able to protect data. It should be scalable, shareable, and manageable. In determining which storage to use, you must consider its accessibility, performance, availability, capacity, security, durability/scalability, and cost.

Depending on what the needs of the business are, a decision must be made on whether the storage will be centralized, distributed or both. In this case, I will choose the storage to be centralized. This is because centralized storage has more efficient storage utilization, has a simpler storage management, and has dedicated storage platforms. It is easy to back up and can restore quickly. Meanwhile, distributed storage’s security and back up is more complex. Distributed storage’s data migration can be a challenge as well, there could be impacts with its performance and capability (having “noisy neighbours”). Although distributed storage is currently cheaper, the future is unknown. It could have a higher long-term cost profile than the centralized storage.

### Evaluation of metadata products and their capabilities

Repository

This is the central file storage location. It can be used to hold multiple versions of files. They are composed of the trunk, which contains the most recent/current version of the project, the branches, which contains the new versions of the program, and the tags, which are used to save the different versions of a project. Repositories can store development files, which is advantageous because this is a more structured way of storing these files and useful with larger projects.

Data dictionary

This is where the names, definitions, and attributes of data are described. This will help in letting the user understand more about the data. It contains detailed information about system components. It is where the data elements, records, programs, systems, files, users, and other system components are documented.

### Metadata Training

There should be a Metadata Training Program for the business’ metadata and business intelligence system. It should include information on what drives the business intelligence system applications (such as data marts, and interfaces that both the internal and external consumers of data use to research, analyze and report and make decisions).

They should include information on the business/company’s business side and systems side of Metadata Management. Under the business side, should be the Primary users (which includes the data stewards), the secondary users, and the executive department of the business. There will be a need to train stakeholders in the business side of metadata so that they will be able to access data to an extent, using it to answer their business and research related questions. Under the systems side, should be the executive department of the business, data administrators, database administrators, the developers of the application and the programmers who maintain it.

The training should be able to let the data consumers/users learn what the relationship is between the metadata and data or the business. Using the metadata system applications and interfaces, the end users should be able to access data (to an extent), and format and analyze it.

### Priorities

* Business Drivers
* Stakeholder Challenges
* Metadata Capability
  + Metadata Strategy
  + Metadata capture and storage
  + Metadata Integration and Publication
  + Metadata Management and Governance

### Business Drivers

* EXTERNAL
  + Digital Self Service
  + Increasing Regulatory Pressures
  + Online Community and Social Media
  + Community Building
* INTERNAL
  + Targeted Marketing
  + Brand Reputation
  + 360 View of Customer
  + Efficient IT

### Sources of data

* + Data modeling tools
  + Existing databases and files
  + Paper documentation
  + Audio files
  + Spreadsheets
  + Structured
  + Unstructured
  + Internal
  + External

### Metadata Performance

Always ensure that the performance of metadata is top notch. This is done through creating a measurement of success and determining what is and is not working. Once this is done, the steps for remediation can be done and check how the business process will need to adapt to create/maintain a high-quality performance of metadata. When the goals of the metadata strategy are achieved, and the metadata is both consistent, integrous, and scalable, there is a very high chance that the metadata performance is high/good.

Some key metadata quality indicators are:

* Completeness
* Accuracy
* Currency/Timeliness
* Consistency
* Accountability
* Integrity
* Privacy
* Usability

### Technology[[2]](#footnote-2)

* + Metadata repositories
  + Data modeling tools
  + Database management systems
  + Data integration tools
  + Business intelligence tools
  + System management tools
  + Object modeling tools
  + Process modeling tools
  + Report generating tools
  + Data quality tools
  + Data development and administration tools
  + Reference and mater data management tools

### Technical architecture for managed metadata environment (MME)[[3]](#footnote-3)

The MME shows the architectural components, the people and processes that are used to gather, retain, and disseminate metadata throughout the business in a proper and systematic way. There are six components in the MME, namely, the Metadata sourcing layer, integration layer, repository, management, data marts, and data delivery layer.

The Metadata sourcing layer is where metadata is extracted from its source and sent into the metadata integration layer or straight through to the metadata repository. The metadata sourcing layer should not be merged with the integration layer (this is a common mistake that will cause a lot of future issues). It should be separated because if they are merged, the sources of metadata will be changed and added, and this will negatively impact the integration process. If they are separated, only the sourcing layer will be affected. This will ensure a safe and clean back up and restart point.

The Metadata integration layer is will take in the multiple sources of metadata and integrates and loads them into the metadata repository.

The Metadata repository is where metadata is gathered, stored, and distributed. It is where metadata is cataloged and stored. There are 4 characteristics of a metadata repository, it should be generic, integrated, current, and historical:

* It should be generic because it should not be application-specific, it should store metadata by metadata field/subject area.
* It should give an integrated perspective of the business’ major metadata fields/subject area. It should contain both current and future metadata.
* It should be updated periodically to reflect the current and future environment (technical and business). It needs to be updated constantly to be considered valuable.
* It should be historical. It should store the historical perspectives of metadata is it changes through time, which will enable the business to understand how they have evolved over time and possibly be able to predict the future of the business through business analysis.

The Metadata management layer systematically manages the metadata repository layer and the other MME components. It can archive, back up, load and query statistics, purge, do database modifications and tuning, environment management, job scheduling, query and report generation, recover, manage security processes, versioning, source mapping and movement, and manages the user interface.

The Metadata mart are usually sourced from a metadata repository. It is a database structure that is made for a group of users who have similar needs. They are useful when a metadata user needs the metadata to be arranged in another way than what is currently in the metadata repository. It is also useful with MME’s with a lot of users. This is because they usually encounter performances issues due to the numerous amounts of table joins that are needed for the metadata reports. Due to the multi-dimensions of the metadata mart, they will not experience this performance issue.

Lastly, the metadata delivery layer, it is the final component of the MME architecture. It is where metadata from the metadata repository is delivered to the end users (including any tools or applications that require metadata). They usually deliver to applications, data warehouses and data marts, both business and technical end users, messaging and transactions, metadata marts, software tools, third parties, and websites and e-commerce.

### Use and effectiveness of the metadata

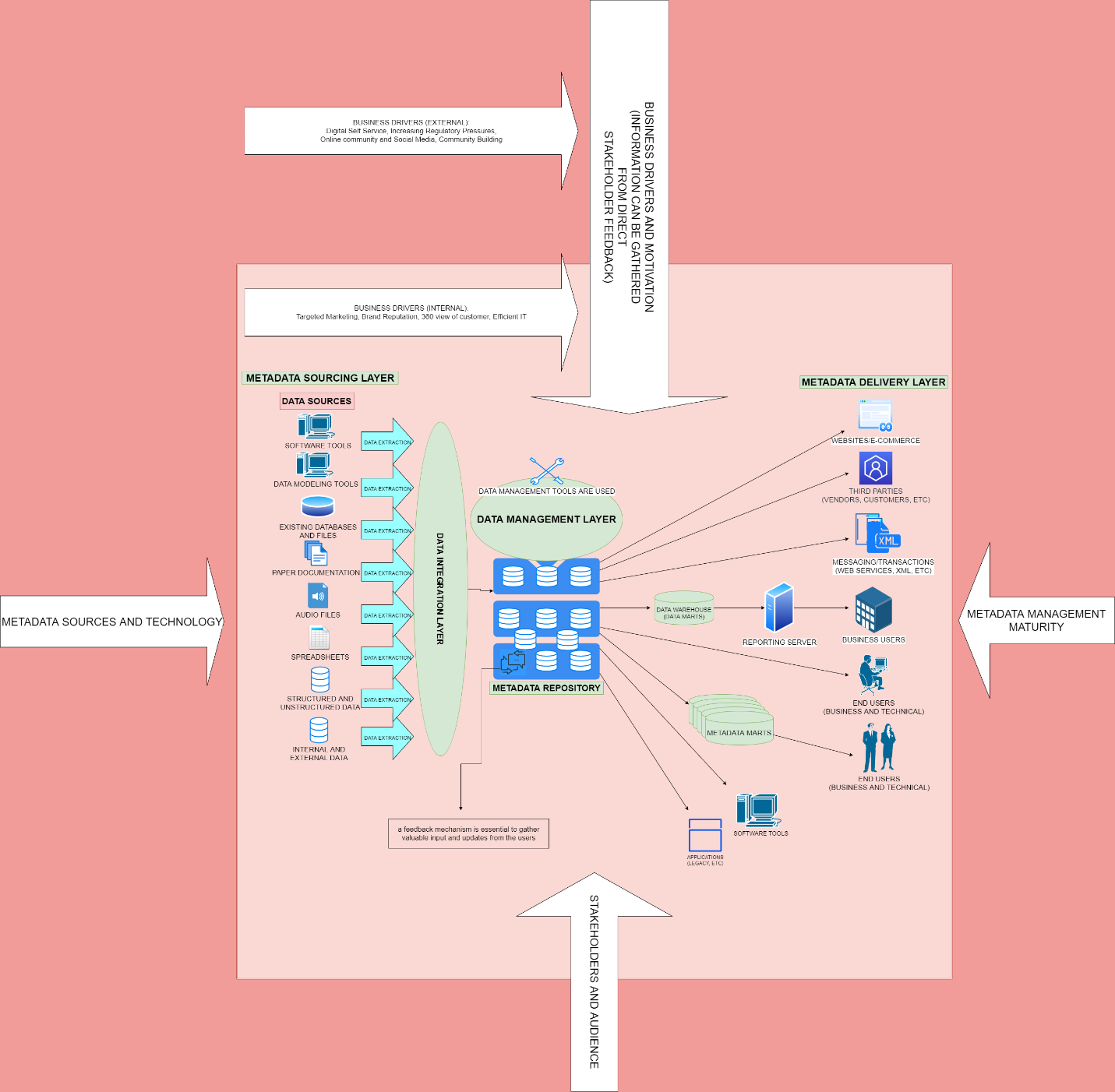
The metadata should be able to provide the data consumers information on what kind of data they have, what the data mean individually and as a whole, where are they located, what transformations they went through, how to retrieve them and how they would be used. Data consumers should also be able to provide feedback to keep the Metadata Management moving forward and getting better. The data consumers should be able to produce meaningful data sets using the metadata.

Stakeholder Feedback

The key business issues and drivers should be determined through direct feedback.

# TASK TWO:

## DIAGRAM OF METADATA STRATEGY

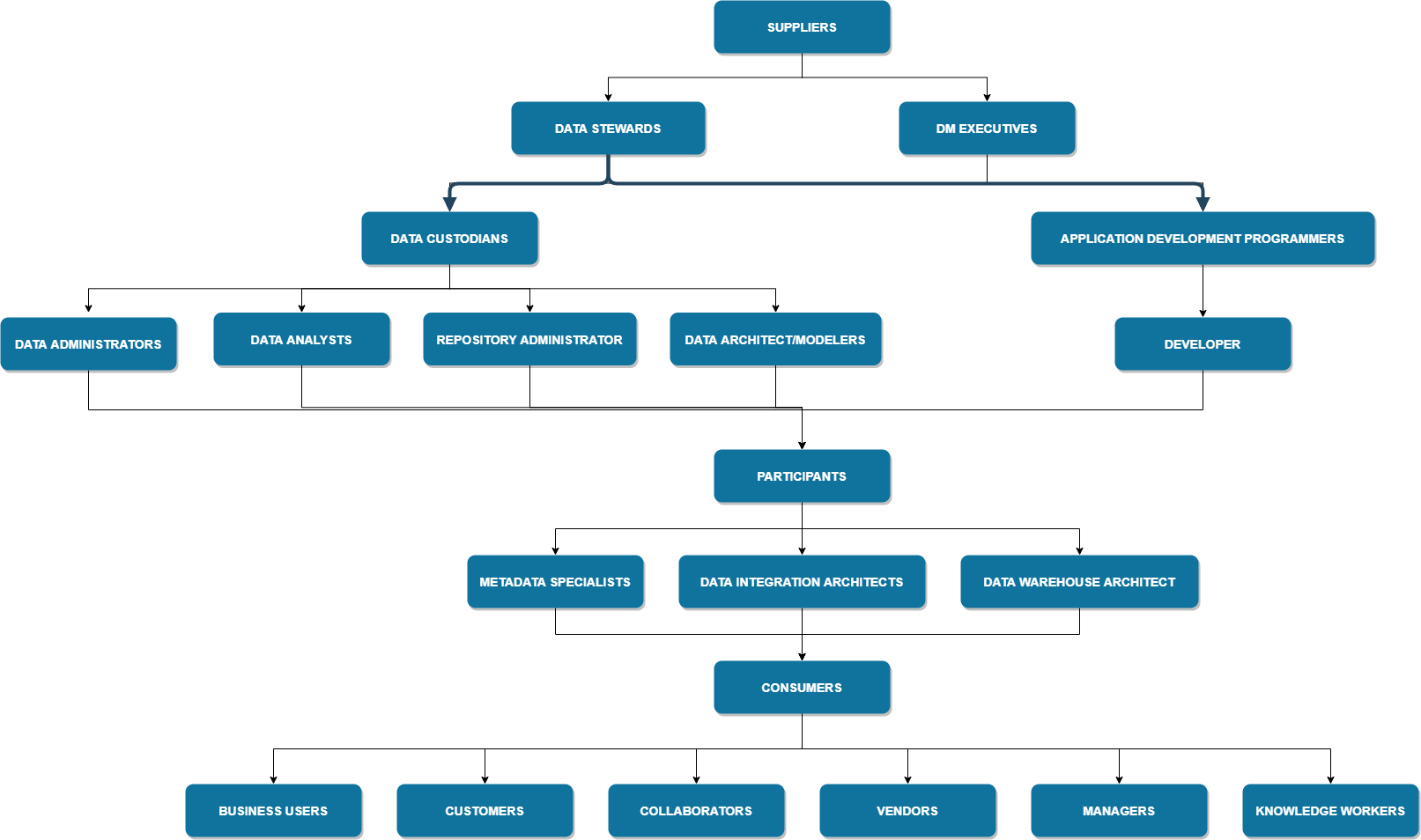


#### DIAGRAM OF METADATA STRATEGY

(PLEASE SEE SEPARATE FILE TO BETTER SEE DIAGRAM)

# TASK THREE:

## ORGANIZATIONAL CHART



#### ORGANIZATIONAL CHART OF PEOPLE INCLUDED IN THE METADATA STRATEGY

(PLEASE SEE SEPARATE FILE TO BETTER SEE CHART)

# References

Accenture. (n.d.). *Metadata Capture: Don’t Miss the Low Hanging Fruit*. Retrieved October 16, 2020, from Accenture: https://www.searchtechnologies.com/blog/metadata-capture-dont-miss-low-hanging-fruit

Aiken, P. (2016, May 10). *Metadata Management Strategies*. Retrieved October 14, 2020, from Slideshare: https://www.slideshare.net/Dataversity/metadata-strategies-61973907

Dufrain. (n.d.). *DATA LINEAGE & METADATA MANAGEMENT*. Retrieved October 17, 2020, from Dufrain: https://www.dufrain.co.uk/data-lineage-and-metadata-management#:~:text=Data%20Lineage%20is%20the%20description,providing%20audit%20and%20compliance%20analysis.

FileHold. (n.d.). *Document Classification and Metadata Capture*. Retrieved October 16, 2020, from FileHold: https://www.filehold.com/features/standard/metadata

Jinks, B. (2017, August 9). *6 Business Drivers for Information Governance*. Retrieved October 19, 2020, from Colligo: https://www.colligo.com/blog/6-business-drivers-for-information-governance/

Kansas Department of Education. (n.d.). *Forum Guide to Metadata*. Retrieved October 15, 2020, from National Center for Education Statistics: https://nces.ed.gov/pubs2009/metadata/appendix\_d.asp

Knight, M. (2017, May 25). *Fundamentals of Metadata Management*. Retrieved October 15, 2020, from Dataversity: https://www.dataversity.net/fundamentals-metadata-management/

Marco, D. P. (n.d.). *MANAGED METADATA ENVIRONMENT (MME): A COMPLETE WALKTHROUGH*. Retrieved October 12, 2020, from EW Solutions: https://www.ewsolutions.com/managed-metadata-environment-mme-complete-walkthrough/

Metz, J., Kim, J., & McDonald, A. (2018, September 11). *Centralized vs. Distributed: A Great Storage Debate*. Retrieved October 14, 2020, from SNIA Ethernet Storage: https://www.snia.org/sites/default/files/ESF/Centralized-vs-Distributed-Storage-Final.pdf

Moddell, M. (n.d.). *The Data Dictionary*. Retrieved October 14, 2020, from Martin Moddell: http://www.martymodell.com/pgsa2/pgsa08.html#:~:text=about%20system%20components.-,Data%20dictionaries%20maintain%20facilities%20to%20document%20data%20elements%2C%20records%2C%20programs,associated%20with%20a%20DBMS%20product.

Park, J.-r. (2009, April). *Metadata Quality in Digital Repositories: A Survey of the Current State of the Art*. Retrieved October 14, 2020, from ResearchGate: https://www.researchgate.net/publication/248920917\_Metadata\_Quality\_in\_Digital\_Repositories\_A\_Survey\_of\_the\_Current\_State\_of\_the\_Art

Schmidt, C. (2019, September 24). *Data Lineage – Basic Elements to Understand*. Retrieved October 17, 2020, from Canto: https://www.canto.com/blog/data-lineage/

Smith, A. M. (n.d.). *Metadata Strategy Overview*. Retrieved October 14, 2020, from EW Solutions: https://www.ewsolutions.com/metadata-strategy-overview/

TechTerms. (2011, August 18). *Repository*. Retrieved October 19, 2020, from TechTerms: https://techterms.com/definition/repository

The Office of the National Coordinator for Health Informatiion Technology. (n.d.). *Metadata Management*. Retrieved October 15, 2020, from The Office of the National Coordinator for Health Informatiion Technology: https://www.healthit.gov/playbook/pddq-framework/data-governance/metadata-management/

Wikipedia. (n.d.). *Metadata*. Retrieved October 13, 2020, from Wikipedia: https://en.wikipedia.org/wiki/Metadata

1. Referenced from: (Knight, 2017) [↑](#footnote-ref-1)
2. Referenced from: (Aiken, 2016) [↑](#footnote-ref-2)
3. Referenced from: (Marco, n.d.) [↑](#footnote-ref-3)