

Rochade - A Metadata Management Technology

Simanta Shekhar Sarmah

Business Intelligence Architect, Alpha Clinical Systems, USA

Abstract Data is extremely powerful phenomena in the current era. The organizations are dealing with huge amount of data in their daily basis. With this kind of information flowing around, it is extremely important to have a proper data governance mechanism in place to manage the data, its components. One of the important aspects of data governance is keeping the information of the data up to date and accurate, link the data to appropriate source and business owners. Metadata management is the way to ensure that information can be accessed, integrated, analysed and used by IT and business to the best effect. There are quite a few tools available in the market for managing the metadata; Rochade is one of them. This paper explores the aspects of metadata management, why it is important for organizations, implementing the lineage, how the mapping is being created in the Rochade tool, and risk analysis through the lineage.

Keywords Metadata, Information Management Systems, Databases, Software Engineering, Repository, Data Warehouse, Data Marts

1. Metadata-At a Glance

Metadata means “data about data”. Metadata simplifies the understanding of the characteristics and usage of data. Metadata is what gives your data a frame of reference. From technical frame of reference, metadata helps IT organizations to manage and maintain data efficiently. From business frame of reference, metadata can be considered as the semantic layer between IT systems and business users.

Forrester Research defines metadata as “the information that describes or provides context for data, content, business processes, services, business rules and policies that support an organization’s information systems.”

Metadata can be of several types in the context of data management environments.

- Technical metadata consists of technical information about the data, such as the name of the source table, the source table column name, and the data type (e.g., data type of a column).
- Business metadata encompasses the business context around data, such as the business terminology, their definition, owners or stewards, and related reference data (e.g., impact analysis using the lineage, risk reports, etc.)
- Operational metadata presents information about the data usage, such last updated date, number of times accessed, last accessed date etc.

2. Need for Metadata Management

Need for effective metadata management is being driven by business initiatives. Data governance includes processes, various policies and standards, organizations, and technologies to manage data like an asset in an enterprise.

The goals of data governance are to:

- Enhance consistency and provide confidence in decision making
- Minimize the risk of regulatory fines
- Improvement of data security

3. Important Terms in Metadata Management

- Data stewardship defines business metadata to link business information, as the business user understands it. Therefore, metadata is very important for data stewards establishing a data governance programs. Metadata management has an important role in data governance by automating the policies that achieve the goals of data accountability and oversight by maintaining an inventory of data and its relationships.
- BCBS (Basel committee on banking supervision) 239 of Jan 2013 makes it mandatory for banks to prioritize and addressing gaps in its Risk Data Aggregation and Reporting (RDAR) capabilities. Senior bank management is unable to obtain an accurate and detailed picture of the risks without these capabilities. In order for banks to assess the risk, data needs to be available in a systematic form. Data is currently available from various sources however it needs to be

* Corresponding author:

sarmah.simanta@gmail.com (Simanta Shekhar Sarmah)

Published online at <http://journal.sapub.org/scit>

Copyright © 2018 The Author(s). Published by Scientific & Academic Publishing

This work is licensed under the Creative Commons Attribution International

License (CC BY). <http://creativecommons.org/licenses/by/4.0/>

extracted and mapped systematically.

4. Introduction to Rochade

Rochade is our enterprise metadata management repository. Rochade is a Meta data tool where the data lineage is created/ updated. There are several reports which the Risk department sends to government or other regulatory department. These reports are analysed and data lineage is created. Data lineage describes as a life cycle of data that includes the origins of data and where it flows over time. It describes what happens to data as it goes through various processes. Lineage Analysis is performed to ensure the quality assurance of the regulatory report is met. Lineage Analysis ensures tracing of the journey of data back to its source.

The main functions of Rochade are as follows:

- Business glossaries at any level – depth, division, enterprise
- Technical metadata describes how data flows between systems – IT or EUC
- Business glossaries linked to technical metadata (link in important reference)
- We can view forward or backward lineage to understand data usage in our IT & business landscape.
- Efficient impact analysis

Following are the basic components of Rochade:

- Content Interface Toolkit (CIT)
- Rochade Web browser

5. Content Interface Toolkit Basics (CIT)

Content Interface Toolkit contains a GUI that enables you to map the contents of the Excel or CSV file to Rochade in a subject area. This toolkit saves the mapping in Rochade for reuse when loading is required in following updates of the file. Typically, the file contents represent an organization's important metadata that should be stored and linked to other assets within the Rochade repository. You can define the data mapping using the GUI, which automatically builds an XML mapping file for the content loader to perform its work.

Content Interface Toolkit includes the following components:

- The GUI to define the mapping of the metadata to Rochade items.
- The content loader to execute the mapping and to populate the repository.
- Implemented as a Web Access Discovery application, Content Interface Toolkit takes either a comma-separated (CSV) or a Microsoft Office Excel (XLS or XLSX) file as input and supports loading specific contents of the file into a Rochade subject area.

Content Interface Toolkit provides these benefits:

- Assists with data warehousing by transforming your metadata into Rochade metadata.
- Enables you to see your entire metadata environment within Rochade.
- Stores data within Rochade information model (RIM) item types.

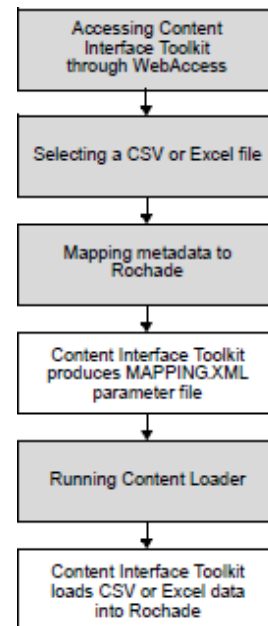


Figure 1. Creation of Lineage

6. Content Interface Toolkit

- Start your web access and point it to Rochade web access by entering this URL: <http://host:port/context> where:

Host is the logical name of the computer where the application server is running.

Port is the number of port where the application server can be reached.

Context is the context path under which you can access Rochade Browser. The default is Rochade, but it could also be the different name.

- Enter your user name and password.



Figure 2

7. Creating New Loader Definitions

To create a new loader definition

- Click the Content Interface Toolkit tab in the Web Access title bar, then the start button on the Content Interface Toolkit start page to create a new loader.

The Content Interface Toolkit main page opens to the Content Interface Toolkit Content Definition dialog:

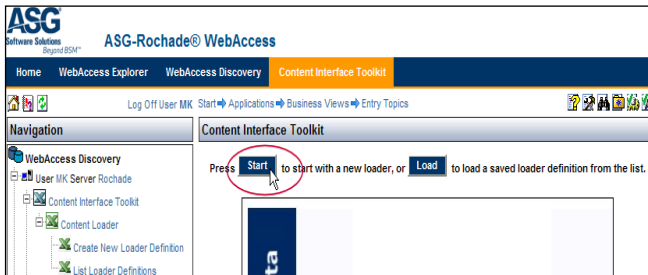


Figure 3

Note: The Save, Discard, and Close buttons in the upper right corner of the dialog enable you to save or discard your changes and close the dialog at any time.

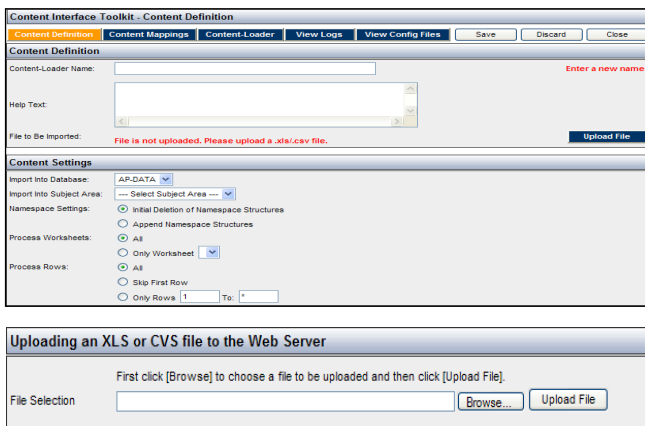


Figure 4

- On the Content Definition panel, specify a name and help text for the loader definition.
- Click the Upload File button. You must upload the CSV or Excel source file to Rochade before you can map its content to the items of a subject area. The Uploading an XLS or CVS file to the Web Server panel displays:

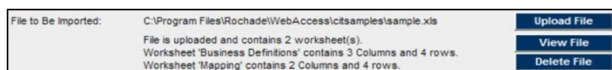


Figure 5

- Enter the path and name of the source file in the File Selection field.
- Click the Upload File button to start the upload of the selected file.
All uploaded source files are stored in Rochade in items of type DSC/SYSTEM in the Web Access control subject area.
- Click the View File button to review the uploaded data on the Content Interface Toolkit - Preview File dialog.

8. Mapping Metadata to Rochade

- Open an existing loader definition or create a new definition.
- On the Content Interface Toolkit - Content Definition dialog, navigate to the Content Settings panel.
- From the Import into Database and Import into Subject Area drop-down lists, select the database and subject area that contain the items to which you want to map the uploaded data.
- In the Namespace Settings section, select one of these options:
 - Initial Deletion of Namespace Structures: Deletes the existing namespace structures of items that are included in the import.
 - Append Namespace Structures: Appends the items to existing namespaces.
- In the Process Worksheets section, select one of these options:
 - All: Processes all worksheets
 - Only Worksheet: Processes only the specified worksheet

In this example, the all option button is selected to process all worksheets. Content Interface Toolkit assumes that the worksheets have a similar structure.

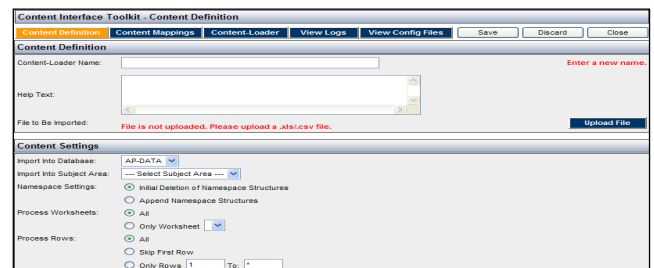


Figure 6

- In the Process Rows section, select one of these options:
 - All: Processes all rows
 - Skip First Row: Processes all but the first row (e.g., if the worksheet has a header)
 - Only Rows xx To yy: Processes only the specified range of rows
- Click the Content Mappings tab to open the Content Interface Toolkit - Content Mapping dialog:

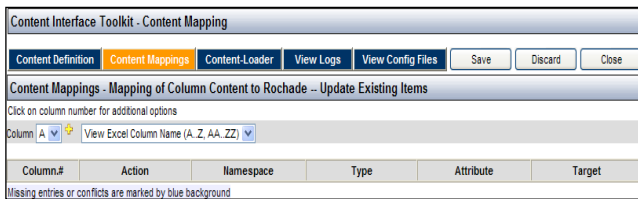


Figure 7

The Column drop-down list displays the columns that are available for mapping. You can list the columns by their name (e.g., the first row of a table), by the Excel column name (e.g., A through Z), or by their sequence number (e.g., 1 to 99).

- Select your preferred view from the drop-down list next to the Add Column icon:
- Select a column from the Column drop-down list, then click the Add Column
- Click the icon next to it to add a mapping for the column:

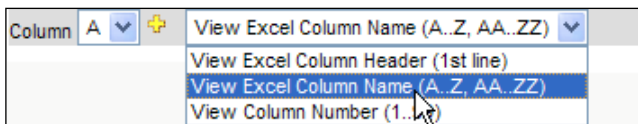


Figure 8

- Select a mapping action for the column from the Action drop-down list.

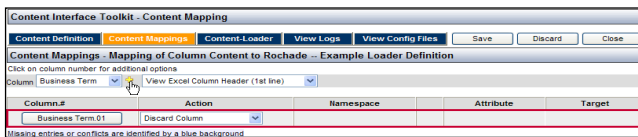


Figure 9

You can define more than one action for a column by adding the column multiple times. The mappings are numbered in the order in which you define them (e.g., for column Business Term, this would-be Business Term.01 to Business Term.nn).

Content Interface Toolkit will load the mapped data into Rochade in the logically necessary order resulting from dependencies between the mapping actions. The load order might not match the order in which you defined the column mappings.

This example uses the first column (Business Term) to create a new Rochade item of item type CIT_ITEM, the second column (Short Description) is mapped to the text attribute Definition, and the third column (Long Description) is discarded.

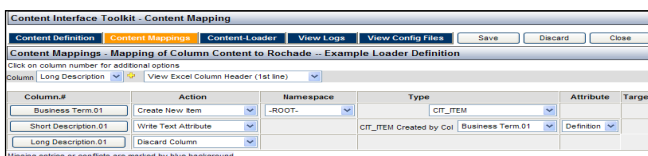


Figure 10

9. Rochade Browser Overview

Rochade Browser includes the functionality to understand and manage metadata.

Metadata Types in Rochade Browser

Rochade Browser operates on the DWRRIM10 information model to provide access to these types of metadata.

- Information on operational data structures used to store them.
- Information on glossary data and business terminology.
- Information on entity relationship(ER) models.
- Information on Extract, Transform, and Load (ETL) processes.

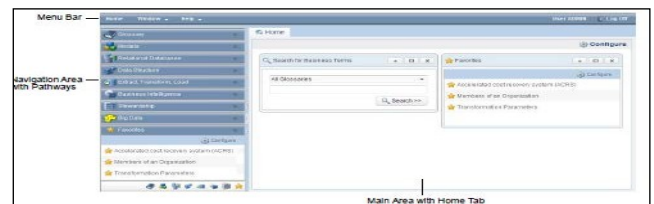


Figure 11

- Information on the contents of data warehouses.
- Information on BI reports and the results of data mining.
- Information on Organizational Entities.

Logging into Rochade Browser

After providing the log id and password, the Rochade Default browser window opens.

Finding Metadata Entities

- Using the Pathway System

This section describes how to use the pathway system to find metadata entities. By following a pathway, you can query the metadata in the Rochade repository and produce lists of corresponding entities. Each pathway has a number of sub- paths that correspond to a specific type of metadata.



The Query Parameter dialog opens:

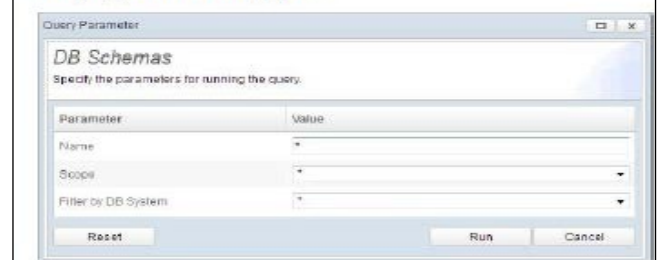


Figure 12

- To expand a pathway to its sub-paths
- Click the title bar of a collapsed pathway to expand it. The Data Structure pathway, for example, has these sub-paths:
- To collapse a pathway
- Click the title bar of an expanded pathway to collapse it.
- To hide a pathway temporarily
- Click the Close button in the title bar of the pathway that you want to hide.

To display the pathway again, click its icon at the bottom of the navigation area:

- To follow a pathway
 - Click the title bar of the pathway that you want to follow to expand it.
 - Click the sub-path that you want to follow (e.g., the DB Schemas sub-path of the Relational Database)

Depending on the pathway you followed, the Query Parameter dialog provides additional, optional query parameters. For example:

- You can restrict the query to entities with a specific name. You can specify the name with wildcards.
- You can restrict the query to a specific scope.
- You can restrict the query to tables or views, respectively.

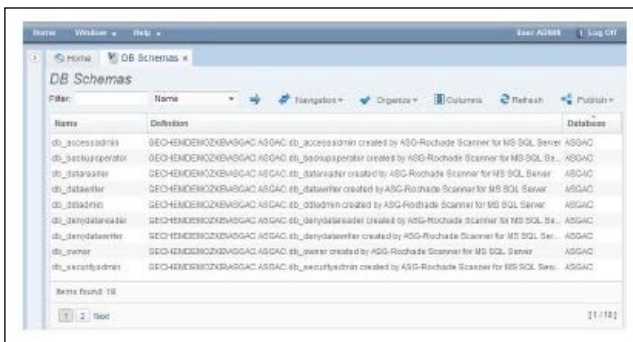


Figure 12

- You can restrict the query to data from a particular database system.
- Click Run to search the Rochade repository for the entities that match your query.

10. Opening Metadata Entities

- To open an entity

Navigate to a list, table, or diagram that contains the entity that you want to open, for example, by using the pathway system or by running a query. Double-click an entity in the result list to open it in tab view.

- Entity Views

Rochade Browser provides two types of entity views—the tab view and the summary view. Entity views display details







about single metadata entities.



Figure 12

The toolbar in entity views provides these controls:

Table 1

Control	Definition
 Open Summary View	Opens the entity in the summary view.
 Add Metanote	Adds a meta note to the current entity.
 Open Tab View	Opens the entity in tab view.
 Navigation ▼	Provides the options for the running queries for the entity.
 Refresh	Reloads the content of the entity view.
 Publish ▼	Provides these options: <ul style="list-style-type: none"> • Send URL: Sends by e-mail a link that points to the entity • Add to Favorites: Adds the entity to your favorites • Exports to PDF: Exports the entity.

11. Running Queries on Metadata Entities

Depending on the entity that you have selected for running the query (i.e., the seed entity), Rochade Browser provides these queries:

- List queries list all entities of a specific type that are related to the seed entity.
- Forward Lineage queries determine the impact if the seed entity changes. They trace the data path (its lineage) downstream from the source. To find the impact, these queries search the references of all related entities.
- Backward Lineage queries determine where an entity comes from. They trace the data path (its lineage) upstream from the source. These paths may run through cubes and database views, ETL processes which load a warehouse or data marts, intermediate staging tables, shells and FTP scripts, and legacy systems.

12. References & Usages Queries

- References search for entities that have direct relationships coming to them from the seed entity.
- Usages search for entities that have direct relationships going from them to the seed entity.
- References & Usages combine the References and Usages queries. They search for entities that have direct relationships coming to them from the seed entity, as well as for entities that have direct relationships going from them to the seed entity.

13. To Run a Query

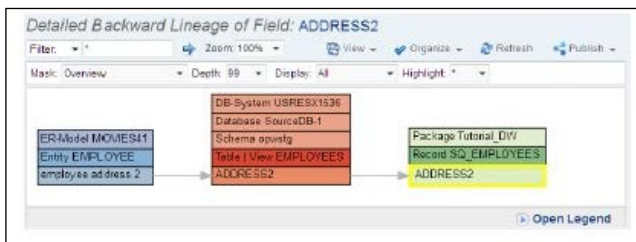


Figure 13

- Navigate to a list, table, or diagram that contains the entity for which you want to run the query, for example, by using the pathway system or by running another query.
- Right-click the entity, then select the query that you want to run from the context menu.

14. Query Result Views

Depending on the query that you run, Rochade Browser displays these different types of query result views:

- Tree views
- Diagram views
- Technical diagram views
- Tree Views

Tree views are the default query result view for Environment and Context queries. Their appearance is similar to table views, with the Name column additionally indicating the relationships between the listed entities.



Figure 14

- Diagram Views

Diagram views are the default query result view for visualizing and analysing the data lineages between entities.

The Diagram View provides the following controls:

Table 2

Control	Definition
Filter: <input type="text" value="*"/>	Selects or filters all nodes that match the specified expression, and refocuses the diagram on the node that is found first.
Zoom: 100%	Represents the diagram at various magnifications.
Mask: Complete Lineage	(Data Lineage queries only.) Enables you to select a mask to show only certain path segments of a query result, and suppress others.
Depth: 5	(Data Lineage queries only.) Sets the depth to which the relationships between the entities are traced.
Display: All	Depicts if all node and relationships in the diagram should be shown (or only currently highlighted ones).
Highlight: LAST_NAME	(Transformation maps only.) Highlights the path to the transformation maps that manipulate the entity from which you started the query

- To select an entity in the diagram

Click the entity's name in the diagram to select it.

- To select a mask for the diagram:

In the Mask drop-down list, select a mask for the diagram:

Based on the data lineage query that you have run, these masks are available:

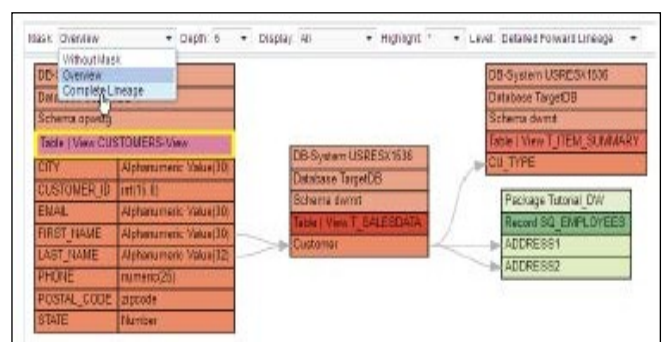


Figure 15

- Without Mask: Shows the full diagram without any restrictions.
- Overview: Shows the lineage between tables and columns, records, etc.
- Complete Lineage: Shows the complete lineage (including transformations)

The information in the diagram is reloaded according to the selected mask:

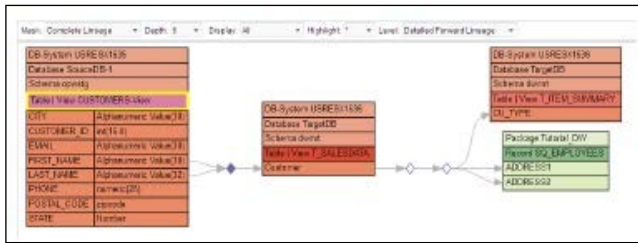


Figure 16

- To highlight specific entities and relationships in a diagram

Right-click the entity whose relationships you want to highlight, then select one of these options from the context menu:

Under this there are two options:

- Highlight References: Highlights the entity, all entities that are referenced by the entity, and the relationships between them.
- Highlight Usages: Highlights the entity, all entities that reference the entity and the relationships between them.

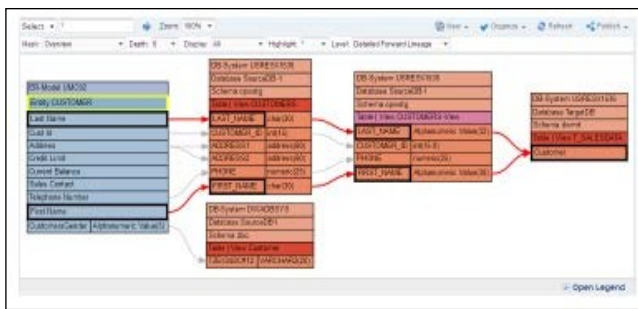


Figure 17

In this diagram, for example, the usages of the entity Customer (on the right end of the diagram) have been highlighted.

- To display only highlighted entities and relationships

From the Display drop-down list, select Highlighted only.

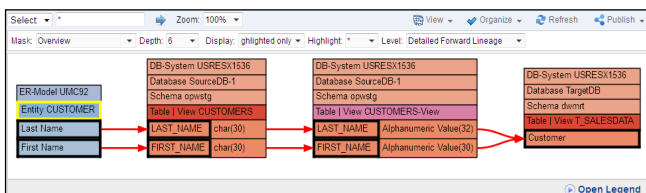
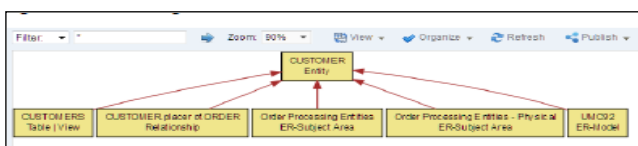


Figure 18

As opposed to the preceding example, the diagram now displays only the highlighted entities, the relationships between them, and the parent entities:

- Technical Diagram Views



They provide a graphical visualization of the relationships between different entities. Technical diagram views provide the same controls as diagram views in addition to display and highlight dropdown lists.

15. Conclusions

Metadata is helpful in understanding the characteristics, usage of data from the basis of data governance. IT is the utmost importance of organizations to have proper mechanism & tools in place for metadata management. Rochade among others like Informatica, Collibra, etc. is a very powerful tool for the metadata management and impact analysis. The metadata management in Rochade is handled through two-stage process-CIT & Rochade explorer. CIT aspects like excel/csv inputs for creating/ uploading data dictionaries, business terms, etc. in Rochade while explorer provides a view of the metadata. The output of exercise enables (business) users to view business terms, its complete lineage from source to end, the transformations logic performed from hopping on from one system to another, enabling business to keep track of the data flowing in & out of the system and making it easier for them to identify the impact of any sort of changes in metadata.

REFERENCES

- [1] Cook, D. (2010). Gold parsing system-a free, multi-programming language, parser. URL: <http://www.goldparser.org>.
- [2] Cui, Y., Widom, J., & Wiener, J. L. (2000). Tracing the lineage of view data in a warehousing environment. *ACM Transactions on Database Systems (TODS)*, 25(2), 179-227.
- [3] Cui, Y., & Widom, J. (2003). Lineage tracing for general data warehouse transformations. *The VLDB Journal— The International Journal on Very Large Data Bases*, 12(1), 41-58.
- [4] de Santana, A. S., & de Carvalho Moura, A. M. (2004). Metadata to support transformations and data & metadata lineage in a warehousing environment. In *Data Warehousing and Knowledge Discovery* (pp. 249-258).
- [5] Springer Berlin Heidelberg. Fan, H., & Poulouvasilis, A. (2003, November). Using AutoMed metadata in data warehousing environments. In *Proceedings of the 6th ACM international workshop on Data warehousing and OLAP* (pp. 86-93).
- [6] ACM. Giorgini, P., Rizzi, S., & Garzetti, M. (2008). GRAnD: A goal-oriented approach to requirement analysis in data warehouses. *Decision Support Systems*, 45(1), 4-21.
- [7] Luberg, A., Tammet, T., & Järvi, P. (2011). Smart City: A Rule-based Tourist Recommendation System. In *Information and Communication Technologies in Tourism 2011* (pp. 51-62).
- [8] Sarmah, S. (2018). Data Migration. [online] Article.sapub.org. Available at:

- <http://article.sapub.org/10.5923.j.scit.20180801.01.html>.
- [9] Missier, P., Belhajjame, K., Zhao, J., Roos, M., & Goble, C. (2008). Data lineage model for Taverna workflows with lightweight annotation requirements. In *Provenance and Annotation of Data and Processes* (pp. 17-30).
 - [10] Springer Berlin Heidelberg. Priebe, T., Reisser, A., & Hoang, D. T. A. (2011). Reinventing the Wheel?! Why Harmonization and Reuse Fail in Complex Data Warehouse Environments and a Proposed Solution to the Problem.
 - [11] Ramesh, B., & Jarke, M. (2001). Toward reference models for requirements traceability. *Software Engineering, IEEE Transactions on*, 27(1), 58-93.
 - [12] Thota, S., 2017. Big Data Quality. *Encyclopedia of Big Data*, pp.1-5.https://link.springer.com/referenceworkentry/10.1007/978-3-319-32001-4_240-1.
 - [13] Reisser, A., & Priebe, T. (2009, August). Utilizing Semantic Web Technologies for Efficient Data Lineage and Impact Analyses in Data Warehouse Environments. In *Database and Expert Systems Application, 2009. DEXA'09. 20th International Workshop on* (pp. 59-63). IEEE.