ELMA EDISON

A00279906

Data Analytics – (AL_KDATA_9_1)

09-11-2020

Role of Data Management

INTRODUCTION

Data Management is the growth, usage and oversight of systems, strategies, projects and practices that track, secure, convey and improve the estimation of information and data resources. It includes some basic activities such as:

Data Policy refers to the general guideline structure that governs the entire data management process. Data Ownership refers to the clear identity of the data owner and decides who has control over data management and subsequent use. Data Documentation is the provision of information required to identify data and the compilation of metadata (data describing other data) relates to dataset summaries that often play a key role in identifying a dataset, for example, Data Time stamps. Data Quality, Standardization and Harmonization, Data Life Cycle Control and Data Audit.

DATA MANAGEMENT PLAN (DMP)

A DMP is an assertion explaining how to preserve and archive the analysis data and where to store the information so as to guarantee its conservation and reuse. The DMP does not need to be long. The National Science Foundation (NSF) expresses that all recommendations must have a beneficial report of close to two pages labelled a data management plan and a restriction of four pages. The plan provides a roadmap documenting the flow of data through the sequential compilation, storage, cleaning, reducing, reviewing, and ultimately archiving stages. In addition, the management plan records the relationships between all the software tools and programs required to direct the data through this study life cycle.

For example, a 2016 DMP includes the principle of FAIR data for the Atlantic Ocean Observation Systems project. The most recent EC-Guidelines updated in July 2016 include more specific recommendations for full life-cycle management through the application of the FAIR principles, which clearly states that the data generated shall be Findable, Accessible, Interoperable and Reusable (FAIR). The application of the principles of FAIR is supposed to be a conceptual integration rather than a technical integration.

A DMP is therefore not a part of evaluation. Optional selection on data management in proposal is worth doing, especially to help justify costs. A DMP is a living or "active" document that should be updated.

DATA QUALITY

As data becomes a core component of any corporate operation, the quality of the data collected, processed and utilized over the span of organization cycles will decide the viability accomplished in working together now and in future. Data quality management includes accuracy, accessibility, completeness and ensures the integrity of information. Clearly, the data is not good, if it is incorrect. Also, if people are unable to access the data due to technical or other problems, the quality of the data will be affected. The key explanation for data inaccuracy established by the research conducted by Experian Data Quality in 2013 was human error, with 59% of the cases evaluated as originating from that source. Avoiding or ultimately fixing low-quality human error-related data needs a systematic approach with the right mix of solutions for individuals, procedures and technology.

Data quality issues drastically effect on achieving key business goals at the corporate level that are inability to respond promptly to new business opportunities, obstacles in the implementation of cost reduction initiatives, shortcomings in meeting increasing requirements for compliance, difficulties in using predictive analysis on the properties of corporate data. These areas, such as Data Profiling, Data Governance, Data Quality Reporting, Data Matching, Customer Data Integration (CDI), Master Data Management (MDM), Digital Asset Management (DAM), Product Information Management (PIM), provide the remedies used to avoid data quality problems and the subsequent data cleaning. The goal of enhancing data quality is to evaluate and enhance a variety of data quality dimensions, such as consistency, uniqueness, relevance, timeliness, accuracy, conformity and completeness.

METADATA

Metadata is a collection of data that explains and gives details on other data. They are pieces of data that have some meaning in relative to another piece of information. Like any other piece of data, it can be developed, handled, preserved and processed. Metadata can be used for anything. The definition of a computer file can be the same as that of a book or a work of art. Both may have, for instance, a title, an author and a year of creation. Metadata adds value to other information but has little value on its own. Text transcripts to audio files is an exception to this rule.

Metadata has to be standardized in order to be useful. This includes agreement on language, spelling, date format, etc. If someone uses a different norm, it can be very difficult to compare data with other data. The schema is the main component of the metadata. Metadata schemas are the overall structure for the metadata. It defines how metadata is set up, and typically discusses standards for common metadata components such as dates, names, and locations. There are also discipline-specific schemas used to address the specific elements required by a discipline.

There is more than one type of metadata. Semantic metadata explains the context in which the data is embedded by describing how it relates to other data that is it goes down to depth explaining the "aboutness" of the data. Metadata on food items, for example, could be listed in detail, including nutritional facts, country of origin, etc. Active metadata is often stored and structured in a manner that allows manual or automated operation of one or more processes. Active metadata makes it simpler and more effective for users to create, deploy and run data management applications for analytics, data science, governance, or almost any other purpose. Not only does it highlight missing, faulty or irregular data, but it can also help to improve the consistency of analytics by quickly fixing and enhancing data to improve decision-making and prevent costly errors. Organizations are able to link, use and discover their data with modern Metadata Management procedures that put semantic and active metadata in place. This makes them more accountable and better able to determine the importance and risks of the data and its use.

REFERENCES

1. Olajide, V., 2019. Introduction to Data Management.
2. Data Quality - What, Why, How, 10 Best Practices & More [WWW Document], n.d. . Enterprise Master Data Management • Profisee. URL https://profisee.com/data-quality-what-why-how-who/ (accessed 11.4.20).
3. Data Quality Management: Framework & Best Practices [WWW Document], n.d. . Study.com. URL https://study.com/academy/lesson/data-quality-management-framework-best-practices.html (accessed 11.4.20).
4. What is Metadata Management? - Turning Your Data into a Strategic Asset, n.d. . PoolParty Semantic Suite. URL https://www.poolparty.biz/what-is-metadata-management/ (accessed 11.4.20).
5. Hayslett, M., n.d. LibGuides: Metadata for Data Management: A Tutorial: Standards/Schema[WWW Document].
   URL https://guides.lib.unc.edu/metadata/standards (accessed 11.4.20).
6. Example DMPs and guidance | DCC [WWW Document], n.d. URL https://www.dcc.ac.uk/resources/data-management-plans/guidance-examples (accessed 11.7.20).
7. https://onlinepubs.trb.org/onlinepubs/nchrp/cd-22/manual/v2chapter2.pdf