Seahorse Project one report

Author: Michael McGrath

Date: Monday, October 1, 2018

Reviewers: TBD

# Version History

V0.0 First draft, not for review

# Table of figures

[Figure 1 Original Data Lifecycle Model 3](#_Toc526154944)

# Introduction

This paper is the output from the first Seahorse project. The goal of this project is to define a working data lifecycle model (DLM) that is practical in use and is as complete as possible. The DLM is based of the Initial data lifecycle model, described below.

## The growing role of data.

The growth in data is exceptional. While estimated of just how much data exists vary all estimates show the same trend, exponential growth. Khoso (Khoso, 2016) estimates that data in 2016 totaled around 4.4 Zetabytes and that that would grow to ten times that by 2020. Even email is growing by over 4% per year (Radicati & Hoang, 2017). In tandem with the growth in data there is a corresponding growth in storage required to persist that data. Additionally we see that the majority of that data is in unstructured forms; Khoso (Khoso, 2016) estimates that about 90% of data created is unstructured, this view is supported by practitioners (V Cloud News, 2015) who also asserted that 90% has been created in the last year.

There is also great value to this data, for example a report by focusing on data in the automotive industry suggests that *“The expected growth of the value pool from car data and shared mobility could add up to more than USD 1.5 trillion by 2030”* (Balasubramanian et al., 2016) and that this data used in AI *“… could potentially deliver additional economic output of around $13 trillion by 2030, boosting global GDP by about 1.2 percent a year”* (“Modeling the global economic impact of AI | McKinsey,” n.d.).

In addition, some practitioners have reported other problems most notably that there are no common data formats, little metadata and little interoperability because the systems exist in independent and distinct systems(Sohn, 2017). It is even asserted that these issues place such a financial burden on the health industry as to be a noteworthy contributor to rising costs (Wang, Kung, & Byrd, 2018).

The DLM is important because it will provide the basis for the next project which will describe an objective method and model to identify the total lifecycle cost and value of data.

## Methodology

The method used is simple. A search of documented DLMs was performed and where there were gaps or differences identified these were evaluated and the original DLM was modified or not, accordingly.

# Initial data lifecycle model

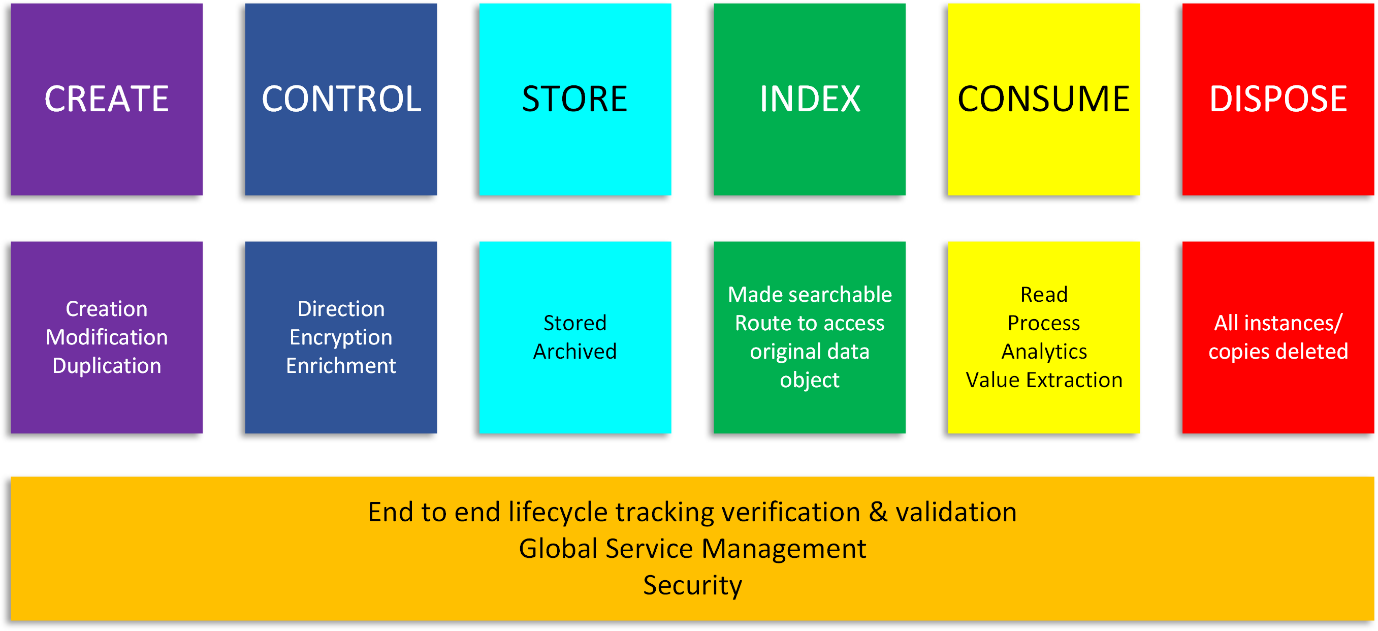


Figure 1 Original Data Lifecycle Model

# Documentation reviewed

# Changes to the data lifecycle model from the literature

# Modified model

# Conclusion

# Appendix

# Bibliography

Balasubramanian, J., Belker, S., Chauhan, S., Colombo, T., Hansson, F., Inampudi, S., … Kässer, M. (2016). Car data: paving the way to value-creating mobility. *McKinsey & Company*, (March). Retrieved from https://www.mckinsey.com/~/media/McKinsey/Industries/Automotive and Assembly/Our Insights/Creating value from car data/Creating value from car data.ashx

Khoso, M. (2016). How Much Data is Produced Every Day? Retrieved August 20, 2017, from http://www.northeastern.edu/levelblog/2016/05/13/how-much-data-produced-every-day/

Modeling the global economic impact of AI | McKinsey. (n.d.). Retrieved September 24, 2018, from https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy

Radicati, S., & Hoang, Q. (2017). *Email Statistics Report, 2017-2021*. *Email Statistics report, 2017-2021 - Executive Summary* (Vol. 44). London. Retrieved from http://www.radicati.com

Sohn, A. (2017). You Want to do What With Your Data? Retrieved September 8, 2018, from https://bigdata.cioreview.com/cxoinsight/you-want-to-do-what-with-your-data-nid-27091-cid-15.html

V Cloud News. (2015). Every day Big Data Statistics - 2.5 Quintillion Bytes of Data Created Daily. Retrieved August 20, 2017, from http://www.vcloudnews.com/every-day-big-data-statistics-2-5-quintillion-bytes-of-data-created-daily/

Wang, Y., Kung, L. A., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, *126*, 3–13. https://doi.org/10.1016/j.techfore.2015.12.019