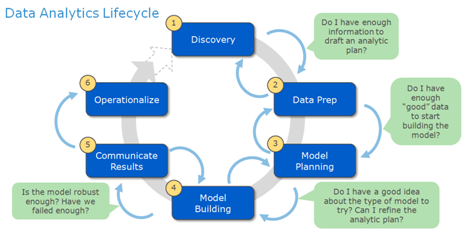
# What is the data analytics Lifecycle?

[**Dr. Michael K Hernandez**](https://mkhernandez.wordpress.com/)

Innovator, Data Analyst, and Diversity Champion, with passions for Weather and Investing



The scientific method helps give a framework for the data analytics lifecycle (Dietrich, 2013). According to Dietrich (2013), it is a cyclical life cycle that has iterative parts in each of its six steps:

* Discovery
* Pre-processing data
* Model planning
* Model building
* Communicate results
* Operationalize

However Erl, Buhler, & Khattak (2016), suggested that it is divided in nine steps:

* Business case evaluation
* Data identification
* Data acquisition & filtering
* Data extraction
* Data validation & cleansing
* Data aggregation & representation
* Data analysis
* Data visualization
* Utilization of analysis results

Prajapati (2013), stated five steps:

* Identifying the problem
* Designing data requirements
* Pre-processing data
* Data analysis
* Data visualizing

Between these three different lifecycle versions, there is a general pattern that emerges, but it also suggests that the field of data analytics is still too nascent to pin down an exact data analytics lifecycle.  For the purpose of this discussion the lifecycle that will be used is from Services (2015), which uses the Dietrich (2013) lifecycle. Note that both Services (2015) and Dietrich (2013) model is iterative and not static steps.  This lifecycle model allows all key team members to conduct planning work up front and towards the end of the data analytics project to drive success (Dietrich, 2013).

## When is it beneficial for stakeholders to be involved?

If following an agile development processes the key stakeholders should be involved in all the lifecycles. That is because the key stakeholders are known as business user, project sponsor, project manager, business intelligence analyst, database administers, data engineer, and data scientist (Services, 2015).  Some of the benefits of applying the Agile development processes to this lifecycle is because it allows for iterative feedback for speed-to-market, improved first-time quality, visibility, risk management, flexibility to pivot when needed, controlling costs, and improved satisfaction through engagement (Waters, 2007).  Allowing the stakeholders to participate in most of these steps can allow the following work to be done to their specifications.

For the first step, discovery, the business learns its domain and its relevant history with lessons learned from previous projects (Services, 2015). Before proceeding ask: “Do I have enough information to draft an analytic plan and share for peer review?” (Dietrich, 2013; Services, 2015). Pre-processing data, also known as data preparation is where a copy of the data is placed in a sandbox (not the original), where the data scientists and team can extract, load and transform (ELT) the copied data (Services, 2015). In this stage, data could also be cleaned, aggregated, augmented, and formatted (Prajapati, 2013). Before proceeding ask: “Do I have enough good quality data to start building the model?” (Dietrich, 2013; Services, 2015). Model planning is when the data scientist and team determines the appropriate models, algorithms, workflow of the data, which helps identify hidden insights between the variables (Services, 2015).  Before proceeding ask: “Do I have a good idea about the type of model to try? Can I refine the analytic plan?” (Dietrich, 2013; Services, 2015). Model building helps sets roughly about 2/3 of the data for training the model and 1/3 of the data for testing the model for production purposes and discovering hidden insights (Prajapati, 2013; Services, 2015). Before proceeding ask: “Is the model robust enough? Have we failed for sure?” (Dietrich, 2013; Services, 2015).   Communicating results could be done visualization of data to the major stakeholders to see if the results are a success or failure (Services, 2015).  Visualization is done in this step is supposed to be interactive with all parties involved in this project (Prajapati, 2013). Finally, the operationalize step is when the data is ready to provide reports, documents, on a pre-defined time interval such that key decision makers could receive the vital data needed (Services, 2015).

## References

Dietrich, D. (2013). The genesis of EMC’s data analytics lifecycle. Retrieved from https://infocus.emc.com/david\_dietrich/the-genesis-of-emcs-data-analytics-lifecycle/

Erl, T., Buhler, P., & Khattak, W. (2016). Big Data Fundamentals: Concepts, Drivers & Techniques. Retrieved from http://www.informit.com/articles/article.aspx?p=2473128&seqNum=11

Services, E. E. (2015). Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data, (1st). [Bookshelf Online].

Prajapati, V. (2013). Understanding the data analytics project life cycle. Retrieved from http://pingax.com/understanding-data-analytics-project-life-cycle/

Waters, K. (2007). 10 good reasons to do agile development. Retrieved from http://www.allaboutagile.com/10-good-reasons-to-do-agile-development/