White Paper: The Air Force Data, Analysis, and Human Analytic Capital Ecosystem

Darryl Ahner, PhD, P.E.
Air Force Institute of Technology

The purpose of this white paper is to identify the key elements needed for the U.S. Air Force to be a data-driven organization and propose an approach for effective use and development of human analytic capital.

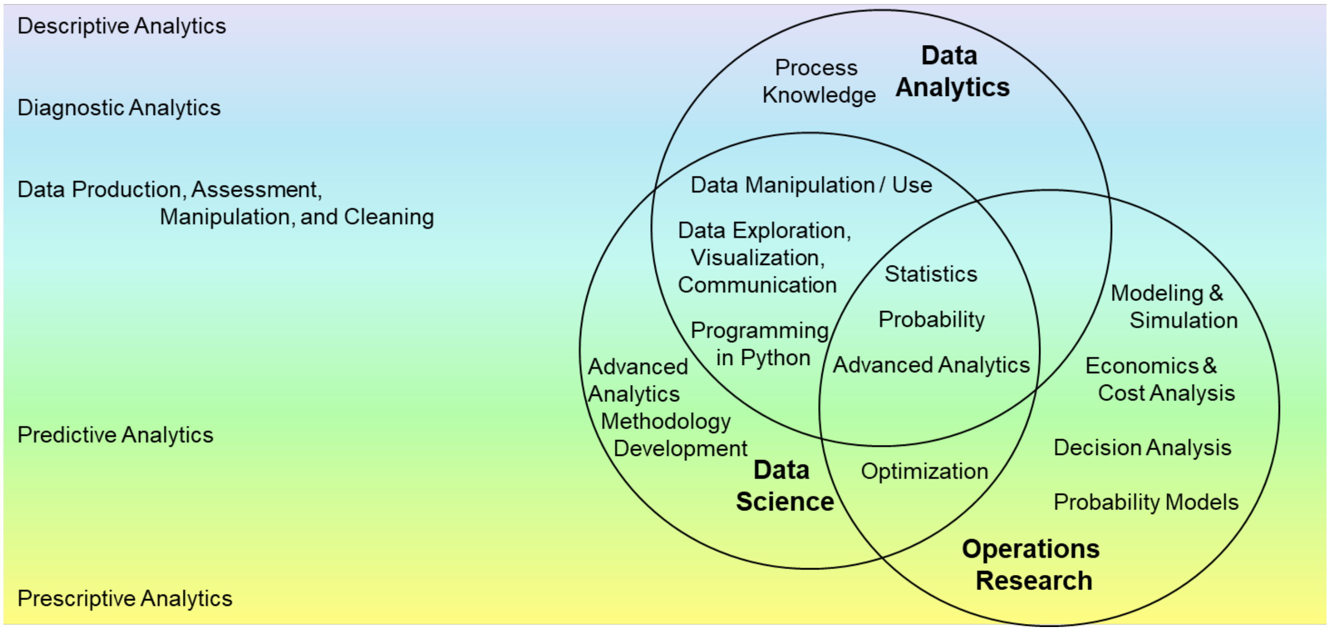
Executive Summary: Data-driven organizations possess certain characteristics that include 1) a leadership culture that requires data analysis to underpin decisions, 2) technology that enables data analytics, 3) treating data as a critical core asset, 4) investing on human analytic capital, and 5) ensuring data access. The presence of specific data analytic process elements are also required to be a data-driven organization which include:

- A Data Collection Strategy
- Data Quality and Structure (volume, velocity, variety, validity, veracity, value, visibility)
- Data Access and Availability
- Process Understanding
- Data Analysis Tools Availability and Access
- Data Visualization, Reporting, and Information Sharing
- Leadership Commitment to include identifying clear goals, building talent, and commitment

None of these characteristics are easily achieved, but perhaps the one with the longest lead time and requiring the greatest emphasis is investing in human analytic capital; none of the specific data analytic process elements are achievable without it. This requires the identification of the talent required and the hiring, education and training development, and talent management of that human analytical capital. One path of acquiring, developing and effectively using human analytic capital is to view the expertise needed through the categories of data analysts, data scientists, and operations research analysts. The Figure below illustrates the interrelationships between these talent pools and types of analytics on which each group focuses. These types of analytics are applied to different types of Air Force analysis required for our Air Force to function most effectively and efficiently, including:

- Resource Allocation Analysis
- Operations & Campaign Analysis
- Weapon Systems Performance Analysis
- Process Analysis for the Air Force Personnel, Logistics, and other systems
- Analytic Intelligence Analysis

A center of excellence focused on assisting data analysts in the field with technical quality control and generation of tools that consists of a key number of data scientists and operations research analysts should be established to further empower quality analysis. Through this approach of identifying characteristics of a data-driven organization, specific data process elements present in such an organization, and an emphasis on a human analytics capital plan, the U.S. Air Force can transform into a data-driven organization and achieve effective use and development of human analytic capital.



Interrelationships of Data Analytics, Data Science, and Operations Research

DISCLAIMER: The views expressed here are those of the author and do not reflect an official Air Force position.

Background: Data-driven organizations possess certain characteristics. These characteristics include 1) a leadership culture that requires data analysis to underpin decisions, 2) technology that enables data analytics such as the cloud, software, algorithms, etc., 3) treating data as a critical core asset, 4) investing on human analytic capital by developing, sourcing, and deploying talent, and 5) ensuring data access. In addition to these characteristics, the presence of specific data analytic process elements are required to be a data-driven organization, the absence of which will undermine effective data analytics. These specific elements to ensure proper functioning of data analytics consist of:

- A Data Collection Strategy
- Data Quality and Structure (volume, velocity, variety, validity, veracity, value, visibility) – This aspect of data involves data finding, cleaning, understanding, and integrating.
- Data Access and Availability
- Process Understanding
- Data Analysis Tools Availability and Access
- Data Visualization, Reporting, and Information Sharing
- Leadership Commitment – This commitment consists of identifying clear goals, building a talent pool, a commitment to data storage and access, talent management.

The Air Force current Information Dominance Flight Plan (2017) is information, not data, focused. It states that the Air Force will establish a Chief Data Office with requisite authority to develop policy, implement cross-enterprise governance structures, grow cross-classification data analytic capabilities and begin to explore what is needed as a first step to becoming a data-driven organization. Ideally, the Air Force Chief Information Officer (CIO), Chief Data Officer (CDO), and Chief Analytic Officer (CAO) would ensure that specific data analytic process elements function in a seamless manner. Figure 1 is a notional model of such a relationship where the CDO ensures proper data collection, quality, structure, and a baseline data analytic capability; the CIO ensures data access and availability, data analysis tools availability, and ensures means to effectively disseminate results; and the CAO provides an extended analytic capability especially where advanced techniques and algorithm development are required. These functions can be thought of as necessary core functions to have a data-driven Air Force that may not be sufficient to achieve quality data from which to make decisions.

Many different types of analysis exist and can be found in various forms in data-driven organizations. They include descriptive, diagnostic, predictive, and prescriptive analytics. Descriptive Analytics includes visualization, trend analysis, summary statistics, Diagnostic Analytics examines data or content to answer the question “Why did it happen?”, Predictive Analytics includes pattern recognition, data relationships, forecasting, and Prescriptive Analytics informs decisions, balances constraints, and optimizes objectives. The CDO baseline analytic capability as proposed in the Air Force Data Analytic Ecosystem would focus on using a preexisting data analytics toolset consisting of the ability to conduct descriptive, diagnostic, and predictive analytics including already-developed machine learning techniques. A CAO extended analytic capability might focus more on prescriptive and predictive analysis. All analysis requires understanding of the process being analyzed.

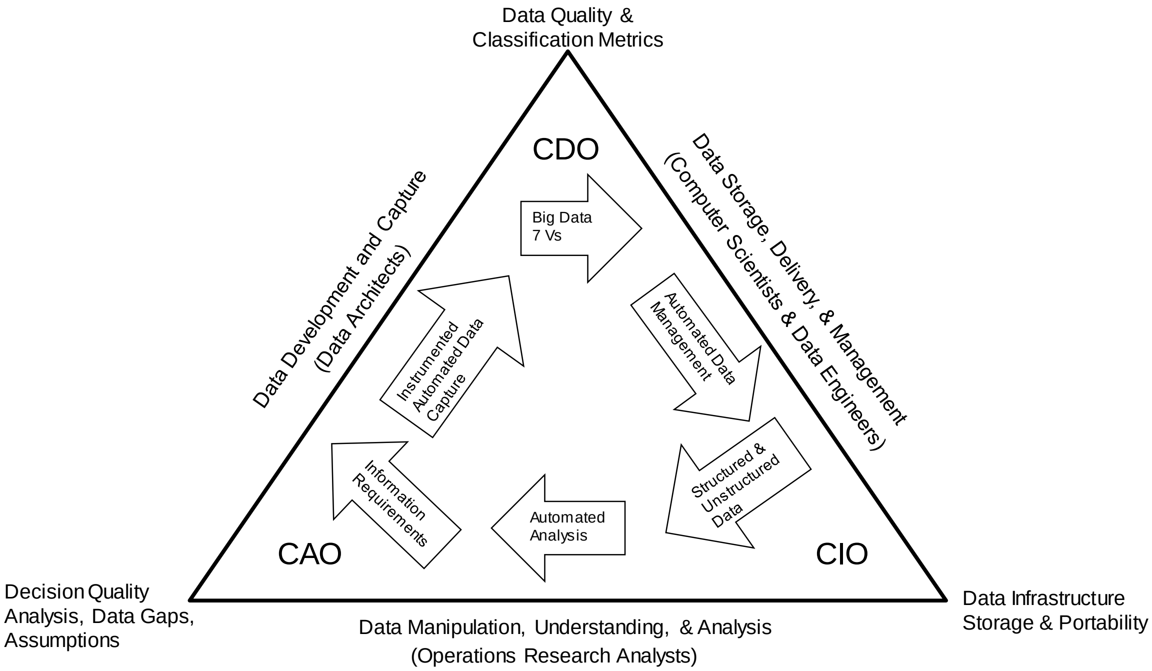


Figure 1 Air Force Data Analytic Ecosystem (Core Functions)

Human analytic talent is essential for proper functioning of a data-driven organization. Some common data analytic positions are provided in the appendix. Data analytics and data science are sometimes used as interchangeable terms; however, there are substantial differences as indicated by significant average pay differences for an experienced data analytics professional ($102,495) as compared to an experienced data scientist ($164,094) in the Washington D.C. area. This is also captured in the following definitions for data analytics and data science.

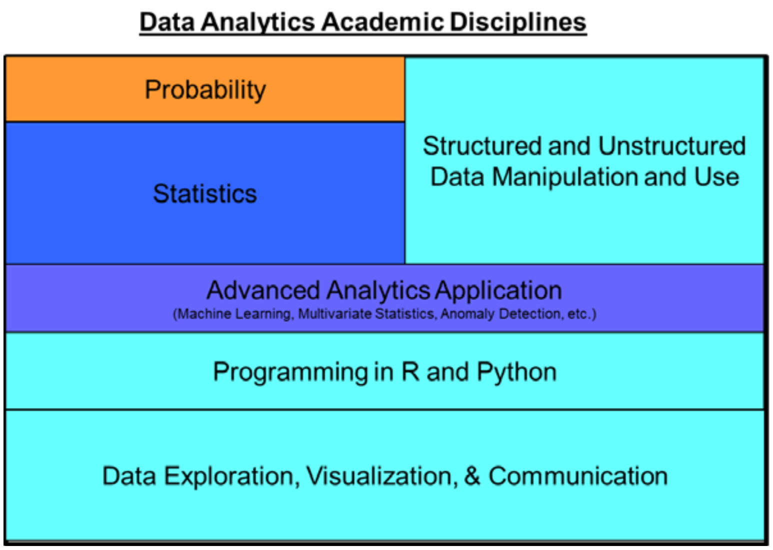
Data Analytics is the process of extracting information from data. It involves multiple stages including establishing a data set; preparing the data for processing; applying statistical, heuristics, or predictive models; identifying key findings and creating reports. The goal of data analysis is to find actionable insights that can inform decision making.

Data Science is a multi-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from data. Data science is a "concept to unify statistics, data analysis, machine learning and their related methods" in order to "understand and analyze actual phenomena" with data. It employs techniques and theories drawn from many fields within the context of mathematics, statistics, computer science, and information science. Data Science is differentiated from Data Analytics by this more rigorous inclusion of theory and technique development.

An Approach for Effective Use and Development of Human Analytic Capital: In addition to requiring computational technology, a data driven organization requires human analytic capital to ensure all data analytic process elements work properly. Data Analyst and Data Scientists are not job titles used in the recent past in the U.S. Air Force. In fact, while civilian sector employment of data analysts and data scientists has grown significantly in the past ten years there has been little attention paid to data analytics in the Air Force in the same period. The Air Force has, however, a deeply rooted Operations Research capability as indicated by the award of the 2017 Institute for Operations Research and the

Management Sciences (INFORMS) Prize for their pioneering and enduring integration of operations research (O.R.) and analytics programs into their organizations.

Data analytics, as defined here, consists of a toolset that depends on several academic disciplines as graphically depicted in Figure 2 with disciplines shared with operations research multi-colored. As the previous definition indicates, its focus is on preparation of data and application of tools to that data. This focus of data analytics lends itself more to an emphasis on descriptive and diagnostic analytics along with data preparation and cleaning. While data analysts may successfully conduct predictive and prescriptive analytics, these analyses may require more in-depth technical understanding than a data analyst normally achieves in a formal degree program.



Data Analytics Emphasis

1. Descriptive Analytics (data visualization, exploration, trend analysis, summary statistics)
2. Diagnostic Analytics (examines data or content to answer the question "Why did it happen?)"
3. Data Production, Assessment, Manipulation, and Cleaning
4. Predictive Analytics (pattern recognition, data relationships, forecasting)
5. Prescriptive Analytics (informs resource decisions, balances constraints, optimizes objectives)

Figure 2: Data Analytics Program Components and Emphasis

Data Science addresses both techniques and theory to extract knowledge and insights from data. As seen in Figure 3, data science consists of all the same academic disciplines as data anaytics, but, instead of just using a toolset, it extends its role to developing new methodologies and processes. Since optimization is required in advanced analytic methodology development, data science disciplines include this discipline as well. Data science places more emphasis on predictive analytics and the advanced techniques and theories needed for novel predictive, diagnostic, and data manipulation activities.

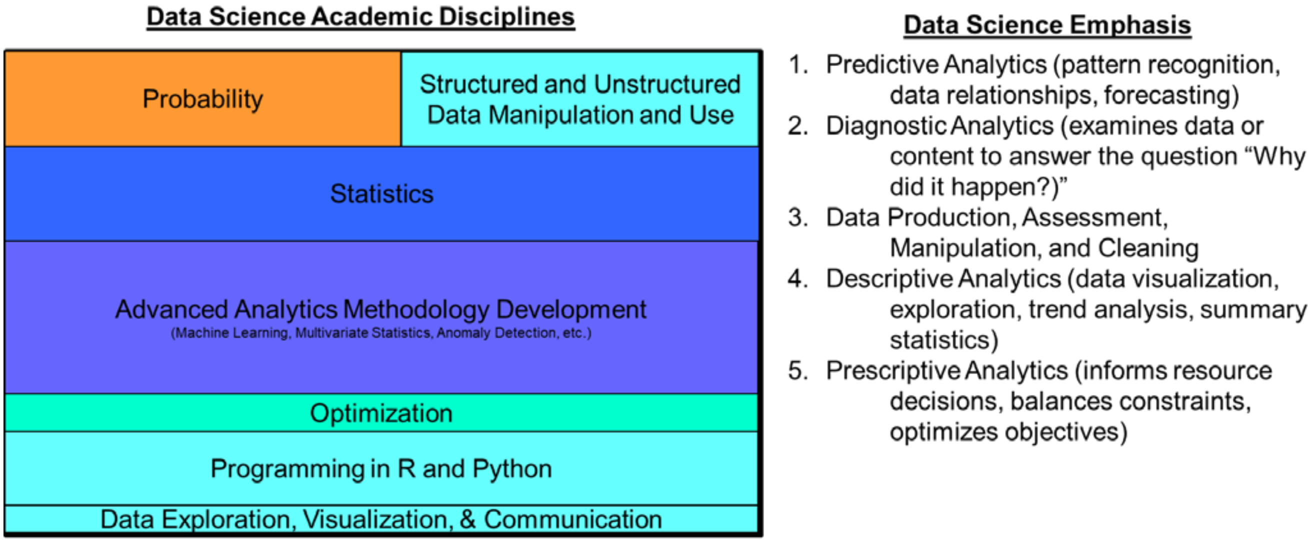


Figure 3: Data Science Program Components and Emphasis

Operations research (OR) has a rich history of informing decisions. As seen in Figure 4, OR academic disciplines overlap with some of the same academic disciplines as data analytics and data science with less of a focus on data and more on use of models whether they are models within simulations, economic, resource allocation optimization, or decision valuation. Typically as far as an academic discipline, the data for use by operations research is assumed to be available. Operations research academic disciplines require a mature understanding of mathematics, to include linear algebra, multivariate calculus, and optimization theory making operations research professionals with advanced degrees a significant analytical resource. As a practical matter, data often is not readily available or useable, resulting in operations research practitioners separately developing skills that align with the data analyst and data science functions.

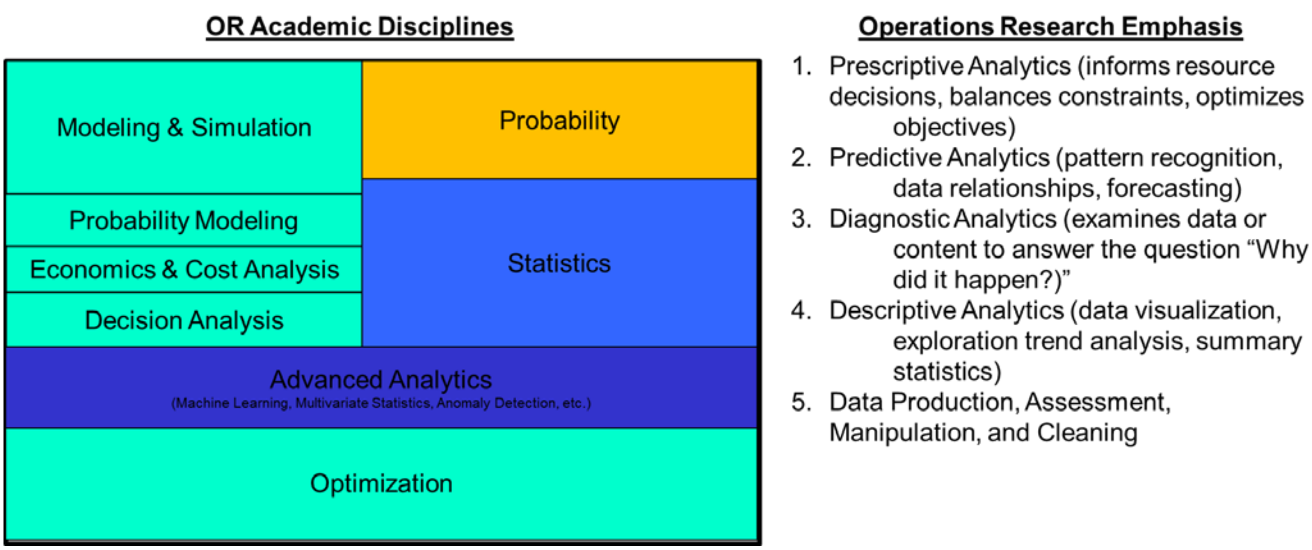


Figure 4: Operations Research Program Components and Emphasis

The question then is how data analytics/science and operations research Air Force functions are similar and how are they different. It is helpful to begin with the types of analysis conducted within the Air Force. These types of Air Force analysis and their relationship to data analytics and operations research are depicted in Figure 5. Data analytics is most applicable to repeated processes where data is readily available where these data can be captured in developed metrics and exhibit stationarity.

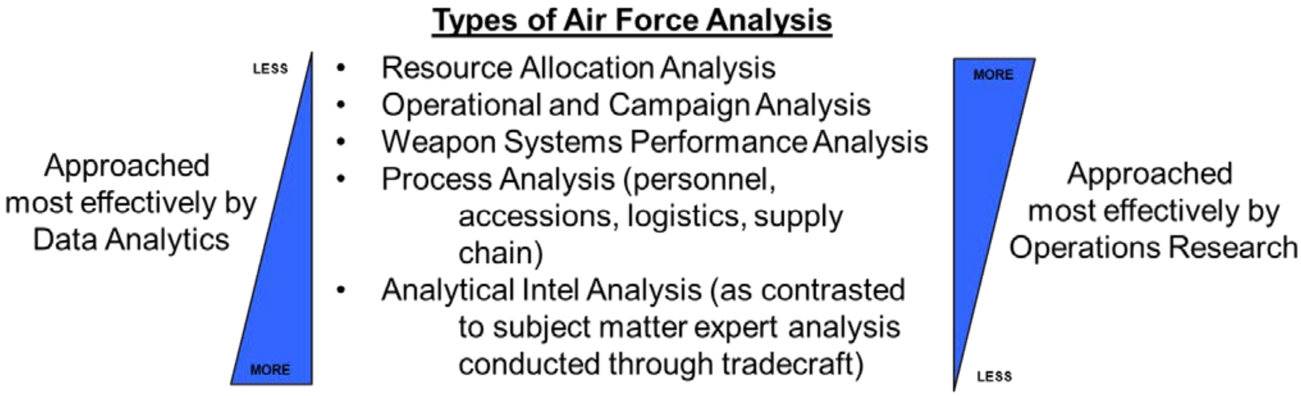


Figure 5: Types of Air Force Analysis Relation to Data Analytics and Operations Research

Operations research is most applicable where data may exist, but this data is not available in large quantities or may not exhibit stationarity, but where the system is understood enough to model it. Stationarity is a rather intuitive concept; it means that the statistical properties of the process do not change over time or space. While there is no hard and fast rule, analyses that address repeatable processes having significant data are better suited for data analytics. On the other hand, resource allocation, operational and campaign analysis, and weapon systems performance analysis do not generally follow this repeatability. Of course, well-educated data analysts, data scientists, and operations research analysts can address analysis of all types, but, their disciplines lend themselves to better accomplishing the previously described types.

While Figure 1 addressed a notional model of such a relationship where the CDO ensures proper data, the CIO ensures data access and results dissemination, and the CAO provides an extended analytic capability, a broader view is required for the U.S. Air Force to become a data-driven organization and have effective use and development of human analytic capital. Operations Research Analyst and Data Scientists support within this notional model but data analytics must take a larger role than just being confined to the inner workings of the Air Force Data Analytic Ecosystem. The Air Force Data Analytic Ecosystem must extend beyond these necessary core functions to a broader data analytics community so that data collection, quality, analysis, and visualization can support decisions at all levels through Airmen from many career fields. Additionally, a center of excellence focused on assisting data analysts in the field with technical quality control and generation of tools would be ideal as a reach-back capability to empower quality analysis. This center of excellence would consist of data scientists and operations research analysts.

With the previous insights in mind, this paper recommends an approach for effective use and development of human analytic capital where operations research analysts continue to be effectively used in their historical capacity while also being leveraged for successful analytics of all kinds, where a focused group of data scientists are developed to address both techniques and theory to extract knowledge and insights from data, and where data analytics for a broader Air Force audience becomes readily available to develop career agnostic data analytics to extract information from data to better inform leaders' decisions at all levels. Focus on and investments in hiring, education and training, and talent management all become critical for this human analytics capital plan to be successful.

In summary, the Air Force requires a leadership culture that requires data analysis to underpin decisions, technology that enables data analytics, treatment of data as a critical core asset, investment in human analytic capital, and data access to achieve the goal of being a data-driven organization. These elements, however, are not sufficient. The presence of specific data analytic process elements are required that include

- A Data Collection Strategy
- Data Quality and Structure (volume, velocity, variety, validity, veracity, value, visibility)
- Data Access and Availability
- Process Understanding
- Data Analysis Tools Availability and Access
- Data Visualization, Reporting, and Information Sharing
- Leadership Commitment to include identifying clear goals, building talent, and commitment.

One path of acquiring, developing and effective use of human analytic capital is to view the expertise needed through the categories of data analysts, data scientists, and operations research analysts while emphasizing hiring, education and training, and talent management. Figure 6 illustrates the interrelationships between these talent pools and types of analytics on which each group focuses. A center of excellence focused on assisting data analysts in the field with technical quality control and generation of tools consisting of data scientists and operations research analyst should be established to further empower quality analysis. Through this approach of identifying characteristics of a data-driven organization, specific data process elements present in such an organization, and a human analytics capital plan, the U.S. Air Force can transform into a data-driven organization and achieve effective use and development of human analytic capital.

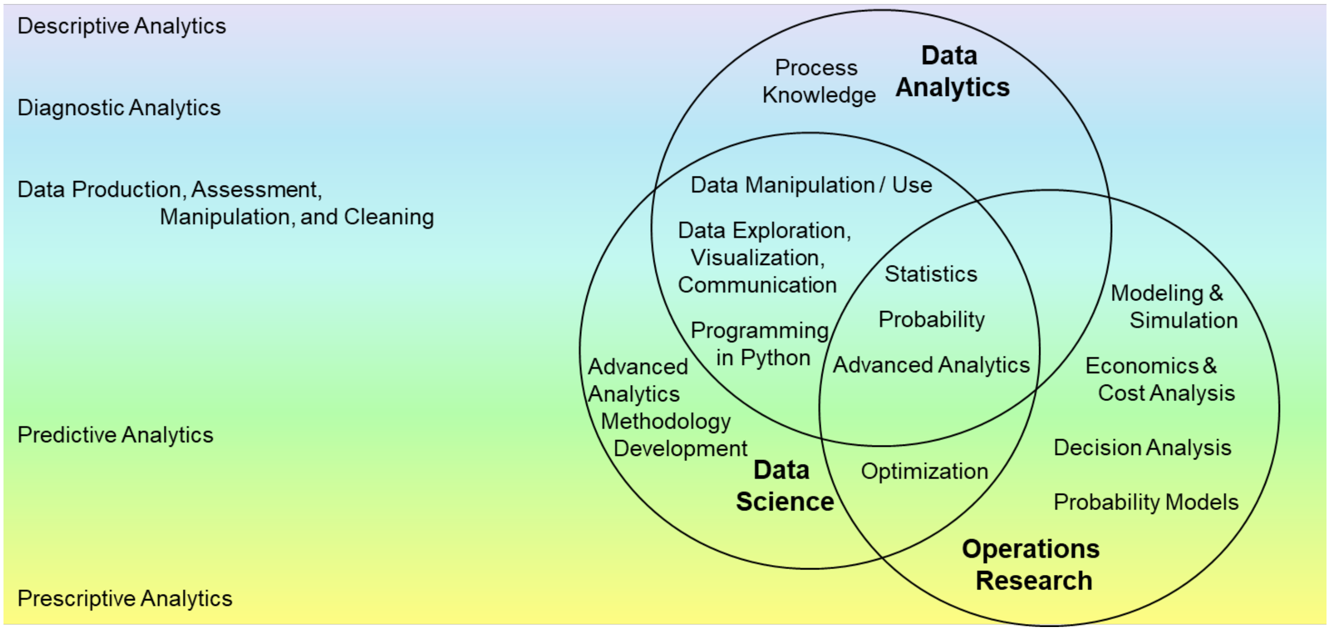


Figure 6: Interrelationships of Data Analytics, Data Science, and Operations Research

Appendix: Common definitions within data analytics and operations research

Data scientists are big data wranglers. They take an enormous mass of messy data points (unstructured and structured) and use their formidable skills in math, statistics and programming to clean, manage and organize them. Then they apply all their analytic powers – industry knowledge, contextual understanding, skepticism of existing assumptions – to uncover hidden solutions to business challenges.

Data engineers build massive reservoirs for big data. They develop, construct, test and maintain architectures such as databases and large-scale data processing systems. Once continuous pipelines are installed to – and from – these huge "pools" of filtered information, data scientists can pull relevant data sets for their analyses.

Data architects create blueprints for data management systems. After assessing a company's potential data sources (internal and external), architects design a plan to integrate, centralize, protect and maintain them. This allows employees to access critical information in the right place, at the right time.

Data analysts collect, process and perform statistical analyses of data. Their skills may not be as advanced as data scientists (e.g. they may not be able to create new algorithms), but their goals are the same – to discover how data can be used to answer questions and solve problems.

Source: https://www.mastersindatascience.org/careers/

Operations research analysts use a wide range of methods, such as forecasting, data mining, and statistical analysis, to examine and interpret data. They must determine the appropriate software packages and understand computer programming languages to design and develop new techniques and models.

Source: https://www.bls.gov/ooh/math/operations-research-analysts.htm

What is Operations Research?

Operations research originated in the efforts of military planners during World War II. More here about what that means:

In the decades after the war, the techniques began to be applied more widely to problems in business, industry and society. Today, operations research is used by virtually every business and government throughout the world and remains an active area of academic research.

Operations Research (O.R.) is a discipline that deals with the application of advanced analytical methods to help make better decisions.

Employing techniques from other mathematical sciences, such as mathematical modeling, statistical analysis, and mathematical optimization, operations research arrives at optimal or near-optimal solutions to complex decision-making problems.

Operations research overlaps with other disciplines, notably industrial engineering and operations management. It is often concerned with determining a maximum (such as profit, performance, or yield) or minimum (such as loss, risk, or cost.)

Operations research encompasses a wide range of problem-solving techniques and methods applied in the pursuit of improved decision-making and efficiency, such as simulation, mathematical optimization, queuing theory, Markov decision processes, economic methods, data analysis, statistics, neural networks, expert systems, and decision analysis. Nearly all of these techniques involve the construction of mathematical models that attempt to describe the system.

Because of the computational and statistical nature of most of these fields, O.R. also has strong ties to computer science. Operations researchers faced with a new problem must determine which of these techniques are most appropriate given the nature of the system, the goals for improvement, and constraints on time and computing power.

Source: https://www.informs.org/Explore/What-is-O.R.-Analytics/What-is-O.R.