DEPUTY CHIEF OF STAFF,

INTELLIGENCE, SURVEILLANCE

AND RECONNAISSANCE

White Paper

DATA SCIENCE *and the* USAF ISR ENTERPRISE

January 2016

DATA SCIENCE AND THE USAF ISR ENTERPRISE

1. Introduction

(U) When today's array of intelligence sources, products, and services are combined with the quickly expanding "Internet of Things" and Big Data technologies, the Air Force Intelligence, Surveillance, and Reconnaissance (AF ISR) Enterprise is presented with both opportunities and challenges. Sifting through and organizing information from multiple sets of dispersed data has long been a hallmark of intelligence analysis; however, what's different today is the exponential growth in potential sources, increasing demands on analysts, and planned changes to Intelligence Community (IC) analysis technologies. Examples abound from across AF ISR of enterprising analysts and collectors employing the latest IC tools to access, visualize, and analyze datasets that previously could not be combined into a single working environment. Although this new generation of tools enable a whole new level of data fusion, **there remains an opportunity to engage data in new ways to deliver deeper understanding.**

(U) The purpose of this paper is to explore the impact of Big Data, the transformation of IC information technology (IT) and how each will drive the necessity for an AF ISR strategic approach to Data Science. This examination is critical as it answers the call of the *USAF Strategic Master Plan* to, "…Increase flexibility and standardization in ISR processes and knowledge" [1] and the 'AF ISR Vision 2023' demand for an "…ISR enterprise that seamlessly ingests data from an even wider expanse of sources, swiftly conducts multi- and all-source analysis, and rapidly delivers decision advantage to war fighters and national decision makers."[2]

(U) Data Science is the key to unlocking Big Data. Big Data, however imprecise the term, is now widely used to refer to data that is too large to reside on a single server and is most often continually flowing.[3] The Air Force, like other IC members, is both a contributor to and consumer of Big Data from the IC Cloud. The IC Cloud is a main feature of the Office of the Director of National Intelligence (ODNI) "IC IT Enterprise" (IC-ITE) program which represents a mass migration of IC data to a common ecosystem. Described by the ODNI, "...IC-ITE moves the IC from an agency-centric IT architecture to a common platform where the Community easily and securely shares information, technology, and resources. By managing and providing the Community's IT infrastructure and services as a single enterprise, the IC will not only be more efficient, but will also establish a powerful platform to deliver more innovative and secure technology to desktops at all levels across the intelligence enterprise."[4] This transformation presents both challenges and opportunities for AF ISR in adopting a Data Science strategy and capitalizing on the wealth of information available from the IC Cloud.

[1] Secretary of the Air Force. "USAF Strategic Master Plan." May, 2015 (http://www.af.mil/Portals/1/documents/Force%20Management/Strategic_Master_Plan.pdf?timestamp=1434024300378)
[2] US Air Force DCS/ISR. "Air Force ISR 2023: Delivering Decision Advantage." Sep, 2014 (http://www.defenseinnovationmarketplace.mil/resources/AF-ISR_2023.pdf)
[3] Davenport, Thomas, *big data@work* (Harvard Business Review Press, 2014), 1.
[4] Office of the Director of National Intelligence. "IC ITE Fact Sheet." (http://www.dni.gov/files/documents/IC%20ITE%20Fact%20Sheet.pdf)

Figure 1: ***Depiction of IC ITE participants and key features***

(U) The very nature of intelligence analysis performed across the IC is undergoing a transformation due in part to rapid technology advancements, but also in large part due to the evolving post-Cold War international environment where individuals and non-state actors are able to organize across state boundaries in order to achieve their goals. Additionally, globally linked economies, transportation, trade, and media assist the spread of influence beyond the traditional state-centric model. Sensitive information like satellite imagery was often the purview of elite intelligence security services; it is now available to individual consumers. This new environment is driving an increased demand from policy makers and commanders for insightful and relevant intelligence knowledge not available in alternative (often unclassified) sources. Adding to this challenge, the characteristics of the intelligence environment since 2000 suggest fundamental change is occurring: an ever-larger volume of data; widening variety (classic intelligence sources, new sensors and types of data, and open sources); increasing velocity (more data and information in motion, every day); and more complex veracity (data duplication, identity, authenticity, and the resolution of each).[5] The ability of Air Force ISR analysts or "Analyst Airmen," to deliver in this new era of intelligence analysis will be predicated in great part on a strategy to shape AF ISR Big Data into a manageable form to meet tactical, operational, and strategic mission needs.

(U) Analyst Airmen are already adapting to the shifting environment to meet mission requirements of the new apparent steady-state crisis and continuous demand for ISR. Examples of Big Data analysis can be found today throughout the AF ISR Enterprise with innovative officers, enlisted, and civilian employees, often supported by industry partners, seizing the opportunities presented by burgeoning cloud activities. However, the following conditions are most common across AF ISR:

1. A growing quantity of data is available for analysts to fuse via a limited set of applications; however, enterprise data is generally not accessed due to lack of integration and connectedness.

2. Inability to dynamically cross-reference and correlate data vertically through organizations and horizontally across like-missions.

[5] US Air Force DCS/ISR. "Revolutionizing AF Intelligence Analysis." Jan, 2014 (http://www.defenseinnovationmarketplace.mil/resources/20140211_IntelligenceAnalysisWhitePaper_PA.pdf)

3. Lack of a streamlined process to coordinate with and disseminate crucial data to other organizations leading up to, during, and after a significant operation, event or crisis.

(U) Today, and to a growing extent in the future, Analyst Airmen may be presented with opportunities to leverage machine learning, data mining, statistics, and algorithms along with traditional analysis tradecraft to deliver insights. However, the Analyst Airman is typically the end-user and benefactor of the processes and products provided by the Data Scientist. This is an important distinction as we consider the use of Data Scientists and required skills needed for future AF operations and fulfillment of the AF ISR 2023 Vision.

(U) Data Science is at the heart of adapting the AF ISR Enterprise to the evolving Big Data environment and a quickly reshaping global intelligence target environment. The following will define and discuss Data Science, particularly as it pertains to the Air Force, and in doing so, examine how Air Force ISR might approach integrating Data Science into its operations.

2. What is "Data Science" and What is a "Data Scientist?"

(U) The term "Data Scientist" only entered the public discourse on Big Data in the past five years and although there is no broad consensus on the definition of "Data Science," it can generally be described as the practice of exploiting and deriving valuable insights from data. This involves extracting knowledge from data sets that may be small, large, structured, multi-modal, multi-lingual, semi-structured or unstructured.

(U) Data Scientists find, interpret and merge rich data sources; ensure consistency of datasets; create visualizations to aid in understanding data; build mathematical models using the data; and present and communicate data insights/findings to specialists, scientists and non-technical audiences. More specifically, a Data Scientist ensures data is ingested, normalized, curated, perpetuated and retrieved in a timely and accurate manner across an enterprise.[6]

(U) Data Science is both a multi-disciplinary study and practice that borrows from several different fields and requires a versatile skill set. Data Scientists primarily hail from the science, technology, engineering, and mathematics (STEM) fields. Depending on the scale of a mission, project or enterprise, Data Scientists may work as individuals or as an aggregate, with each team member focusing on a particular aspect of the data at hand. Many organizations, both in industry and government agencies, are widening the aperture of Data Science functions to include both technical proficiency and operational/subject matter expertise by way of this team approach and include professionals from the following disciplines:

- Computer Science
  - — Data mining
  - — Machine learning
  - — Computer programming

[6] LaPonsie, Maryalene. "Data Scientists: the Hottest Job You Haven't Heard Of." August 10, 2011. (http://jobs.aol.com/articles/2011/08/10/data-scientist-the-hottest-job-you-havent-heard-of/)

- Mathematics
- Operations Research
  - — Quantitative analytics
  - — Statistics
  - — Regression and Correlation
- Information Theory

(U) Even so, it is important to note that Data Science is different than data analytics. A data analyst—much like an intelligence analyst—looks for gaps in data and collection methods and seeks to fill them, or assesses the data to make projections about future events. Similarly, data analysts determine which data is most useful to customers and which reporting mechanisms are easiest to view and understand. In contrast, the Data Scientist will take that direction and 1) mine large and disparate data pools to discover new relevant data, 2) transform new data types into forms suitable for fusion and analysis (e.g. through agile coding), and 3) statistically model data and uncover correlations between data sets.[8]

(U) Consider this example: AF ISR sensors collect a great deal of data on adversary air activity in a given target area and, as sometimes happens, a new IC data source enables further understanding of the activity. The Data Scientist might develop an application for combining the data sets in a way not otherwise possible with separately existing datasets. It is the intelligence analyst who then uses this application to assess the activity and offer predictive insights. While Analyst Airmen certainly work alongside Data Scientists to achieve common goals, the data analytical function is very different from the data exploitation/"science" function performed by the Data Scientist.

4. Data Science in AF ISR and the DoD

(U) Although Data Science as a distinctive Air Force career field does not exist, some roles lend themselves to the purview of a Data Scientist. For the Air Force, the 61A and civilian 1515 series, Operations Research and Analysis, is gaining the most attention. The Air Staff A9 is already examining ways to build Data Science experience.

(U) Currently the 14N/1N0/1N1/1N2/1N3/ and 1N4 career field education and training plans do not include concerted attention to the basic tenets of Big Data analysis. Exposing Analyst Airmen to the larger Big Data picture could provide a better understanding of key Data Science concepts and the role of Data Scientists, knowledge managers and fusion analysts.

(U) The U.S. Army, Navy and Marine Corps share a similar experience. Although some Military Occupational Specialties (MOSs) include the work of Data Science, there are no Data Science-specific military occupations in any of the services. However, all branches are exploring various ways in which Data Science can assist in intelligence and operations. In June 2015, the Office of Naval Research hosted a Focus Area Forum in which members discussed how both technology and manpower can contribute to advancing Data Science and enhance real-time operational decision making.[7]

[7] *Ibid.*

(U) There is also an increased demand for U.S. Government personnel in a Data Scientist role. An August 2015 ODNI job advertisement for a Data Scientist supports the common characteristics and capabilities described above[8]:

1. Exceptional graphics and/or data visualization skills.

2. Extensive knowledge of appropriate analytic methods and methodological tools in one or more of the following areas: Applied Mathematics (e.g. probability and statistics, formal modeling, computational social sciences); Computer Programming (e.g. programming languages, math/statistics packages, computer science, machine learning, scientific computing); and Visualization (e.g. GIS/geospatial analysis, telemetry analysis).

3. Extensive knowledge of individual, organizational, and technical or transnational issues of national security concern.

4. Subject matter expertise or familiarity in (Counter)-intelligence and/or Security.

(U) Across the IC and military branches, calls for increased Data Science and Intelligence Analysis integration, along with improved technical proficiencies on the part of Data Scientists, are on the rise. This trend seems to be an acknowledgment of the limitation of assigning non-technical personnel to perform Data Science work inside IT enterprises. Increasingly, the military is accepting the professional differences and similarities among Big Data analysis, Data Science, intelligence analysis, knowledge management, and traditional operations.[9]

5. How Data Science Enables AF ISR

(U) As the AF ISR community integrates into IC-ITE, Joint Information Environment (JIE), Defense Intelligence Information Environment (DI2E), and simultaneously maintains its own large enterprises that collect, exploit, and disseminate data, the Data Science discipline and the need for imbedded talent will become more important. Technological advances in live data streaming and correlation allow for real-time decision making on a scale never before experienced in AF ISR. We now have the ability to ingest disparate data sets, put relevant conditions and rules in place, and derive insights and prescriptive intelligence in an unprecedented fashion (see Figure 2)[10].

Figure 2: ***Disparate Data Types***

[8] Sample Office of the Director of National Intelligence Data Scientist position advertisement: https://www.usajobs.gov/GetJob/ViewDetails/406476900

[9] Interviews with Government and Industry leaders and contributors to IC and AF ISR technology projects, 1 June – 30 June 2015.

[10] *Ibid.*

(U) Across the IC, there is a compelling need for Data Scientists to extrapolate data insights into user-friendly formats for further analysis, such as converting raw data to target entities in support of Object-based production (OBP). As data is tagged and becomes "discoverable," additional insights can be gleaned by aggregating data sets against each other, or just the reverse—drilling down to the metadata level to discover new insights. The end result can be a report, visualization or another type of product that may continue to feed data into the overall enterprise and/or contribute to a decision point. An example of this process is provided in Figure 3.

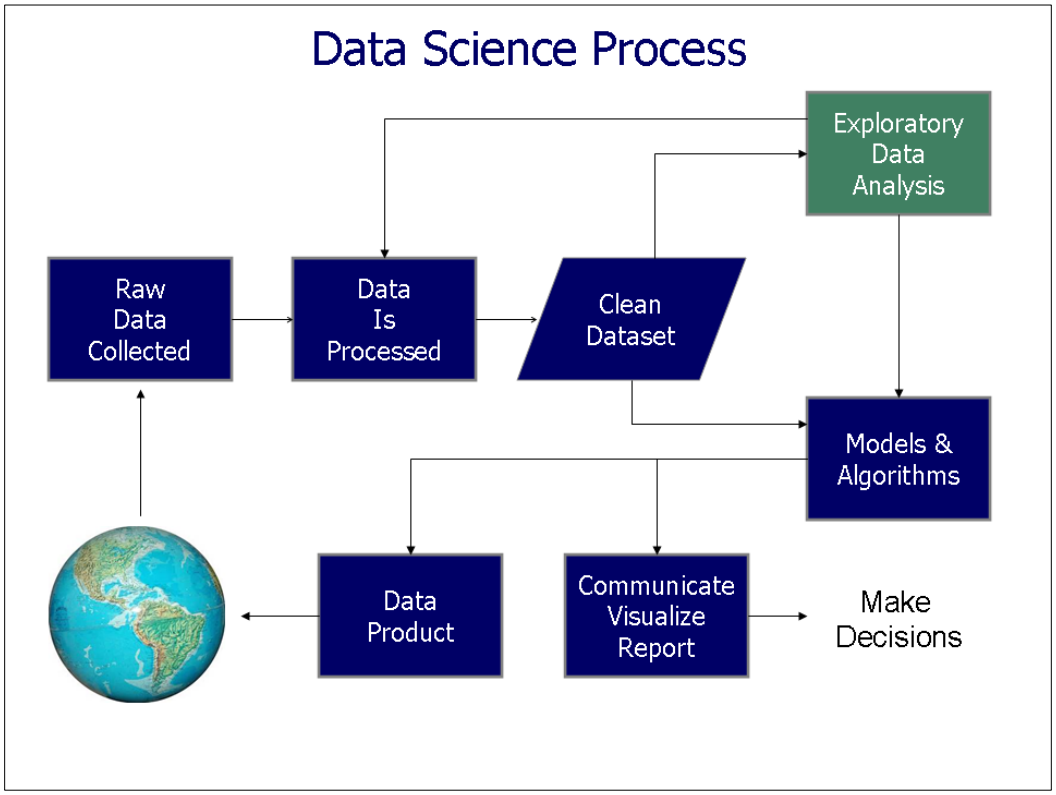


Figure 3: ***Data flow in the Data Science Process***[11]

Diagram shows the Data Science process. Data is collected from sensors in the environment, represented by the globe. Data is then "cleaned" and otherwise processed to produce a data set (typically a data table) usable for processing. Exploratory data analysis and statistical modeling may then be performed, primarily by the Data Scientist with input from data analysts. A "data product" is presented in the prescribed format for data customers and consumers. The product(s) can also create data and feedback into the environment.

(U) Future Data Scientists will operate in a dynamic operational environment with thousands if not tens of thousands of potential data streams. The desired end state is a move from the "Common Operational Picture" (COP) model—where visualization is prized—to a "Common Operational Data" (COD) environment in which data across an entire enterprise is dynamically ingested, correlated and routed. Data Scientists will enable these data-driven architectures that will provide unprecedented access to high volumes of data in real/near real time, enhancing AF ISR and operational capabilities.

[11] O'Neil, Cathy and Rachel Schutt. *Doing Data Science: Straight Talk from the Frontline.* O'Reilly Media, Inc: 2014.

(U) As data-driven architectures are created and implemented, Analyst Airmen can truly take advantage of Big Data analytics on a larger scale. The January 2015 "Five Examples of Big Data Analytics and the Future of Air Force ISR" white paper provides the operational impact of Data Scientists:

- *"Imagine in the future that a PACAF air operations center (AOC) analyst is examining air activity in the South China sea over the previous two weeks, and notes a pattern of flights from select PRC bases to outpost in the Paracel and Spratley Island groupings. Using an app, the analyst isolates bases of origin and destination, and filters the past four months of data to visualize the activity. She discovers a pattern which may be a shuttle operation of troops to outposts from which the troops apparently do not return to home base. This activity is then reported by the analyst as a previously unknown buildup of PRC forces in disputed islands which may lead to international confrontation.* ***Our ability to discover this kind of activity today is severely restricted by an inability to understand what we have already got. The data is derived from varying sensors, compiled in separate databases, and not accessible and manipulable by any single application.****"*[12]

  At the most basic level, when data is ingested and tagged appropriately, it is much easier and faster to retrieve subjects/entities of interest. As Data Scientists organize, curate and aggregate data sets, information can be rapidly and cohesively discoverable and exploited across an enterprise. When analysts detect anomalies in activity or in particular areas of interest, he or she can efficiently filter those same entities and factors to determine if there is a pattern of events unfolding.

- *"Imagine in the future a military strike against a terrorist target in a Central Asian nation using an RPA (remotely piloted aerial vehicle). Our commanders want to know the success of the strike. An analyst, drawing on near-real-time imagery and past information about the site and activity around it, uses an application which detects all changes. In addition, the application provides a visualization of the reactions of both people and object in the target vicinity. Synthesizing this information rapidly, the analyst can provide near-real-time Battle Damage Assessment to the commander, reporting that the primary physical target was destroyed, that bodies were present, and that vehicles appeared to take some persons away from the target area at speed. Although communications from the high value individual (HVI) ceased at the strike, the vehicle departure with a body is included in the assessment that 'the target was physically destroyed, X persons killed and Y possibly injured, therefore we are confident the HVI was injured or killed in the action.' At another level, the theater commander is appraised in near-real-time of the results of several simultaneous strikes, providing an assessment of campaign effectiveness.* ***Our ability to execute both kinds of assessments today is hampered by lack of access to multiple sources, varying levels of security controls, a lack of tools to rapidly correlate and visualize the data, and lack of command and control applications to aggregate***

[12] Deputy Chief of Staff, Intelligence, Surveillance and Reconnaissance. *Five Examples of Big Data Analytics and the Future of Air Force ISR* White Paper, Jan, 2015.

the reports into a near-real-time campaign battle damage assessment.*"*[13]

Another major benefit of Data Science is the ability to correlate data across the "INTs" and previously stove-piped systems. No longer hampered by stand-alone systems, analysts will be able to dynamically correlate real/near real time data in an All Source and Multi-INT fashion. If an analyst needs to provide an assessment on the success of a strike, he or she can cross-cue OSINT, GEOINT, SIGINT and HUMINT updates and reporting to assess the outcome more efficiently.

- *"Imagine in the future a EUCOM analyst is tasked to look at an incident of civil unrest in southeastern Lithuania. After composing and executing her query and defining an area of interest, the system presents not only information on the events in question, but also that a fellow analyst is looking at a similar event in Estonia that two other events of the past week are under examination by others. Examining the project folders of these analysts, she then follows a thread about Russian troop movements along the borders and an aerial reconnaissance intercept of a NATO platform by a Russian fighter. Collaborating with these and other analysts, an intelligence estimate is produced that projects a building confrontation of Lithuanian and Estonian separatists with host countries, and potential provocation by Russian border elements.* ***This type of assessment is difficult to produce today as data and information sets are often segregated by type of source and by regional assignment. In addition, while analysts can collaborate today, it is more often a "pull" system where one asks those who are known to be working a problem, rather than a "push" system where analysts may be automatically alerted to other, similar work****."*[14]

  COD architectures grant not just the rapid retrieval of historical data but for live data as well. Data Scientists can set conditions of interest up front and "train" data sets to seek out analysts directly. This alleviates the requirement to manually search for information and frees up that time for other analysis work. Imagine if data were not merely poured into a database when ingested into an enterprise, but also dynamically routed to individuals who need it!

6. A Strategic Approach to Data Science for AF ISR

(U) For all of the technical and organizational aspects of Data Science, cloud computing and OBP, the human component remains central to all. Activity-based intelligence (ABI) is an emerging form of analysis that seeks to draw meaning from new and more diverse types of data to help determine common behaviors, relationships, and intentions. The AF ISR Enterprise must continue to examine opportunities in these three areas and plan for an evolution toward making each the norm rather than the exception. The AF ISR community must also consider the below investments, enhancements, and challenges for the future.

(U) 1. UNDERSTAND BIG DATA OPPPORTUNITIES**:** Any approach to Data Science and Big Data requires significant forethought. Data Science capabilities are limited and expensive and must

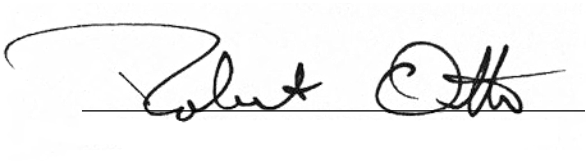
[13] *Ibid.*
[14] *Ibid.*

be applied to the right mission in the appropriate manner. Further study of this area is needed to understand the impacts of leveraging Data Scientists across the AF ISR enterprise. A Big Data specialist specifically tasked with reviewing and researching Data Science methods and approaches may be the key in advancing Data Science in the AF. The individual would:

1. Inventory and identify ISR data resources from a holistic perspective
2. Identify mission sets with potential for improved data-based products and services
3. Integrate AF ISR Big Data efforts with those of Defense Intelligence and the IC
4. Identify strategic partners both internal and external to the AF
5. Identify strategic AF ISR locations to place ISR Data Stewards/Offices
6. Establish an AF ISR Big Data strategy

(U) 2. CONTINUE TO LEVERAGE DATA SCIENCE EXPERTISE**:** Since Data Science is an inherently technical field, it is best performed by trained, qualified individuals with academic and job experience commensurate with the complex and specialized nature of the work. Thus, adding "Data Science" to an intelligence analyst's job description would both diminish the focus on his or her core competency (intelligence analysis) and also result in sub-optimal Data Science. Indeed, it will pose a challenge for government and military personnel to keep pace with rapidly changing data sources, systems, technology, and methodologies. A near-term solution, therefore, is not to require our current force to become Data Scientists, but rather smartly contract for, develop internally, and/or leverage IC Data Scientists to both add operational capability and define and develop training and education objectives to familiarize the force with this emerging and important field.

(U) **CONCLUSION:** With a measured approach to develop a Big Data strategy and incorporate Data Science into AF ISR, we are preparing our Analyst Airman for a future where available data, intelligence consumer demands, and analysis production requirements will all continue to grow. Our goal should be to develop a force of ISR professionals, operators, and knowledge managers that understand and wield the capabilities that Data Scientists bring to the fight. Though AF intelligence Analyst Airmen may never be truly "fluent" in Data Science, we must seek a "conversational" grasp of its principles, processes, and applications. As we further enter the era of IC-ITE, the IC Cloud and emerging data-based analytic techniques such as OBP and ABI, these skills will become increasingly important. **Moreover, each analyst must continue looking for new ways to harness the massive amounts of data so as to produce the best decision-advantage AF ISR can offer.**



ROBERT P. OTTO, Lt Gen, USAF
Deputy Chief of Staff, Intelligence,
Surveillance, and Reconnaissance

OPRs: Mr. Jon "Doc" Kimminau, Analysis Mission Tech Advisor, AF/A2D (jon.a.kimminau.civ@mail.mil)

Col Brian J. Ray, AF/A2DA (brian.j.ray2.mil@mail.mil)