Open Sources For the Information Age

Or: How I Learned To Stop Worrying and Love Unclassified Data

By

James M. Davitch

Individual Research Paper

4 April 2016

Instructor: Col Jeffrey Donnithorne

Blue Horizons Fellowship, Center for Strategy and Technology (CSAT)
Maxwell AFB, AL

DISCLAIMER

The views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the Air Force, the Department of Defense, or the U.S. Government.

Biography

Major James Davitch is currently a Blue Horizons Fellow at the Center for Strategy and Technology, Air University, Maxwell AFB Alabama. Prior to this assignment, he was the Intelligence Officer Functional Manager at Headquarters United States Air Force at the Pentagon. Major Davitch was commissioned in 2002 and his previous duties have included various intelligence officer positions at wing and AOC levels. He is a 2007 graduate of the United States Air Force Weapons School where he was also an Intelligence Weapons Instructor following his initial Weapons Officer assignment at the 3rd Wing, Elmendorf Air Force Base, AK. He has deployed in support of OPERATION ENDURING FREEDOM and OPERATION IRAQI FREEDOM as the Chief of Unit Support at the United States Air Force Central Command in Qatar. His education includes a Bachelor of Arts degree in International Politics from Penn State University, and a Master of Arts degree in International Relations from the University of Oklahoma.

Part I

Introduction

After years of major spending on intelligence, surveillance, and reconnaissance (ISR) collection platforms and sensors, the Air Force is beginning to make a commensurate investment in technology to improve intelligence analysis.[1,2] However, absent a change that recognizes the increasing value of open source Big Data and promotes critical thinking to counter its fixation on classified information and dubious production goals, the Air Force will not realize a return on its investments. The Air Force faces challenges in the Information Age as it seeks to capitalize on emerging technology. But these challenges present opportunities for much needed progress in intelligence analysis.

This paper is divided into four main parts. The first will address why the U.S. Intelligence Community (IC) emphasizes classified information over open source in the execution of its mission and its implications on organizational effectiveness. The paper will trace the origins of the modern IC while also describing the self-reinforcing negative feedback loop that seeks reliance on classified data to the exclusion of openly available information. Part II describes the expanding digital universe of open source Big Data and the technological requirement for a solution as outlined in the Secretary of Defense's Third Offset concept. The paper will present this solution as necessary, yet insufficient, without corresponding increases in critical and creative thinking to capitalize on the technology's promise. Part III will investigate the utility of open source information. It will provide real world examples of how open source Big Data is opening new avenues for intelligence professionals while identifying additional limitations of the prevailing infatuation with secrets. This section concludes with a new way of thinking about open source information as the key component for future early indications and warning (I&W)

about potential adversaries' intentions. Part IV presents a recommended way forward for the AF ISR Enterprise and possible steps for implementation of an open source-focused Department of Defense (DoD) IC.

Harnessing the analytic potential in open source data, rather than closely guarded secret information, is ***the*** Big Data challenge facing intelligence professionals. In spite of that, the IC bureaucracy remains locked in habitual patterns focused narrowly on classified sources. Analysts, focusing on these sources to describe "what happened," often fail to conduct predictive analysis that might prevent surprise. The first step to understanding why this is the case requires observation of the organizational culture of the IC and the barriers to creative thinking it engenders.

A Culture of Secrets

Building intelligence as a communal effort began shortly after the end of World War II. The attack on Pearl Harbor and the succeeding post-war menace of Soviet communism provided the impetus for the development of an intelligence bureaucracy historically unprecedented in size.[3] The combination of the Pearl Harbor attack's surprise and a decentralized intelligence apparatus divided between the U.S. Army and Navy led many to believe the U.S. was vulnerable to another unforeseen attack. This led President Truman to consolidate the intelligence mission, which, he hoped, would identify preliminary I&W of foreign aggression. Thus the 1947 creation of the Central Intelligence Agency (CIA) was, in essence, a hedge against future surprises.[4] So began the period of Industrial Age intelligence running from 1947 to about 1990.

The IC's narrow focus on the development and capabilities of the Soviet Union made "national intelligence" the primary feature of collection and reporting. The Soviet Union represented a complicated target and information about it was sparse, relative to today's internet-

connected environment. But as an intelligence problem it was "comparatively less complex" to today's globalized, interconnected, and interdependent geopolitical setting.[5] Then, intelligence analysts were quick to evaluate any and all bits of available information and prized the scant pieces of mostly restricted data. That data was purported to have predictive power once fitted together with other adjoining pieces of data. Intelligence analysts believed that if the necessary pieces to the intelligence puzzle were found, they could form a more accurate composite picture.[6] The Soviet Union's closed society and impressive counterintelligence architecture made necessary the development of expensive sensors and platforms to provide highly sought after puzzle pieces in this denied environment.[7] Dr. Gregory Treverton described at some length the rationale behind the methods employed during the early years of the IC. He wrote, "In the circumstances of the high Cold War, there were powerful arguments for targeting intelligence tightly on the Soviet Union, for giving pride of place to secrets, especially those collected by satellites and other technical means, and for centralizing intelligence."[8] Furthermore, while the U.S. military had requirements or intelligence needs, the primary customer for Cold War intelligence was the president and National Security Council. After all, it was the president who would bear the brunt of the blame if the U.S. were surprise attacked again. Fortunately, by virtue of its Industrial Age setting, the Soviet Union's mechanistic structures and hierarchical organization provided U.S. analysts markers of military mobilization that were relatively easy to discern. Much of the strategic I&W process during this time focused on "bean counting" Soviet aircraft, ships, and other military equipment. Sensors and platforms were tailor-made to deliver answers to these types of I&W problems. When the IC looked to open sources, it observed mostly the official messages and propaganda sent from the Soviet high command to the masses; this, it was presumed, might provide insight into the leadership's thinking.[9]

The IC's singular focus and exclusive consumer base dictated "a certain logic to the way intelligence was – and is – organized."[10] The National Security Agency (NSA) carried out collection and analysis of signals intelligence (SIGINT), the CIA's Directorate of Operations provided human intelligence (HUMINT), and the National Geospatial-Intelligence Agency (NGA)[11] focused on geospatial intelligence (GEOINT). These organizations existed to funnel primarily secret information up to national-level decision makers. At their inception, they were designed with one logical raison d'etre; to specialize in areas where they could "each concentrate on the distinct contribution [they] would make to understanding the Soviet Union."[12] That understanding was supposed to translate into strategic I&W that would enable national-level leadership to prevent surprise. Specialization begat "INT" stovepipes as each "became formidable baronies in their own right"[13] and continue to this day. These stovepipes perpetuated a failure to think creatively about problems outside of any organization's particular "lane" and, as the Defense Intelligence Agency's Chief Analytic Methodologist wrote, "impose traditionally distinct and narrow perspectives."[14] If one's toolbox contains only SIGINT hammers, then problems might only look like SIGINT nails.

The national intelligence architecture and its business methods stressing the preeminence of secret information imprinted themselves on the DoD intelligence process. Warfighting organizations became intelligence factories, specializing in their particular brand, hammering out products built from classified data like an assembly line. Even today, thousands of civilian and uniformed workers are wedded to a Cold War methodology of producing classified material. For example, imagery sensors produce raw data, which is turned into finished imagery products. This GEOINT production process is a staple of MQ-1/9, Global Hawk, and U-2 missions. The same is done in a similar manner for SIGINT data and the rest of the INTs. Open Source Intelligence

(OSINT) is often relegated to providing supplementary information, as Joint Publication 2-0 suggests.[15]

One reason the IC clings strictly to classified sources is based on the way it responds to the collection priorities laid out in the National Intelligence Priorities Framework (NIPF). The contents of the NIPF are classified, but its purpose is to provide senior policy officials a vehicle to dictate a prioritized list of "critical interest"[16] issues to the IC. The document does not specify using a specific sensor versus a specific collection target and the IC does not necessarily interpret these collection requirements as best fulfilled by classified collection capabilities. But the IC's toolkit is filled mostly with instruments that produce classified data. Therefore, it attempts to address NIPF priorities with the resources at hand and the methodologies deemed "proven" by victory in the Cold War. The NIPF is a prioritized matrix of categories used to "guide and inform decisions concerning the allocation of collection and analytic resources."[17] This prioritization scheme, in turn, influences the DoD's allocation plan of national intelligence resources and tactical-level employment of those forces.[18] This perpetuates the acquisition and development of new sensors and new platforms for the production of more classified information. While the NIPF provides necessary guidance from policy-makers to the IC, it is the IC's method of responding to intelligence problems by looking predominantly to classified sources that merits review.

One problem with this process is that it results in poor metrics for determining ISR effectiveness. Civilian and military collection managers prioritize their collection requirements, derived from the NIPF at the national level and command-driven intelligence requirements below that, based on the priority of the intelligence needed to support a given mission.[19] Quantitative measures, such as the number of missions tasked, and the number of images

collected and processed are often used as proxy measurements to evaluate the effectiveness of meeting prioritized collection requirements. These data are easy to numerically collect and aggregate, but it distracts from investigating qualitative indicators that might determine whether or not the intelligence process is contributing meaningfully to solving the underlying intelligence problem.[20] It also speaks to a failure to both critically analyze problems and devise creative solutions.

For military intelligence personnel, the entire process of collection management as defined in Joint Pub 2.0 hinges on the best use of high demand/low density assets. The Joint Pub is riddled with warnings as to the value of using relatively inexpensive openly available unclassified sources to meet intelligence requirements. It labels open source information, "susceptible to manipulation and deception," and, "subject to source bias and inaccuracy," as if it were the only intelligence discipline liable to these hazards.[21, 22] Nowhere do doctrinal military publications address open sources' growing ubiquity or their potential to respond to priorities outlined in the NIPF.

Anthony Olcott writes that, as a model of how intelligence is to be gathered and used, the intelligence collection process "remains unchanged from the earliest days of the CIA, when information was understood to be in short supply."[23] To break free from this mindset, the intelligence community requires critical and creative thinking about intelligence problems, and at least tacit acknowledgement, if not outright doctrinal documentation, that open sources properly synthesized can be as value-added as other classified means.

Breaking the current paradigm is difficult, but essential, if the IC is to assume a more proactive posture. Barriers to this goal include organizational inertia, the fear of untested alternative methods, and the satisfaction of answering simpler questions, no matter how illusory

their utility. Graham Allison and Philip Zelikow, describing the behavior of large organizations, noted they seldom respond to change until after a crisis and instead follow established routines and simple standard operating procedures.[24] In this vein, Robert Jervis wrote, "If a decision-maker believes that a policy is better than the alternatives on all relevant dimensions, he will react very slowly to evidence that it is failing to reach some of his goals because he will believe that it is still best on other dimensions."[25] Under the prevailing intelligence collection construct, intelligence professionals perpetuate organizational inertia by engaging only in what Ronald Garst defines as descriptive analysis.[26] For example, analytical cells routinely provide statements describing what happened, when, and where in a descriptive manner, eschewing predictive analysis.[27] Not coincidentally, the U.S.'s intelligence sensors excel at providing data that supports descriptive intelligence analysis. But to this end, the IC is reactionary and fails to address what decisions makers are often more interested in -- describing what *will happen* and why. Daniel Kahneman refers to this tendency as the "substitution heuristic"[28] whereby one simplifies difficult tasks by evaluating a related, easier, question.

The reliance on secret data derived from classified sources lends itself to answering questions of whether a collection asset was tasked, whether it collected information, and whether that information was processed. Substituting the tasked-collected-processed metric for the more cognitively difficult predictive analysis allows intelligence personnel to avoid addressing whether the collection effort answered the question it was tasked to answer. So the question, "Is our collection posture working to learn more about the enemy?" becomes instead, "How many ISR sorties have we flown in support of the commander's priorities?"

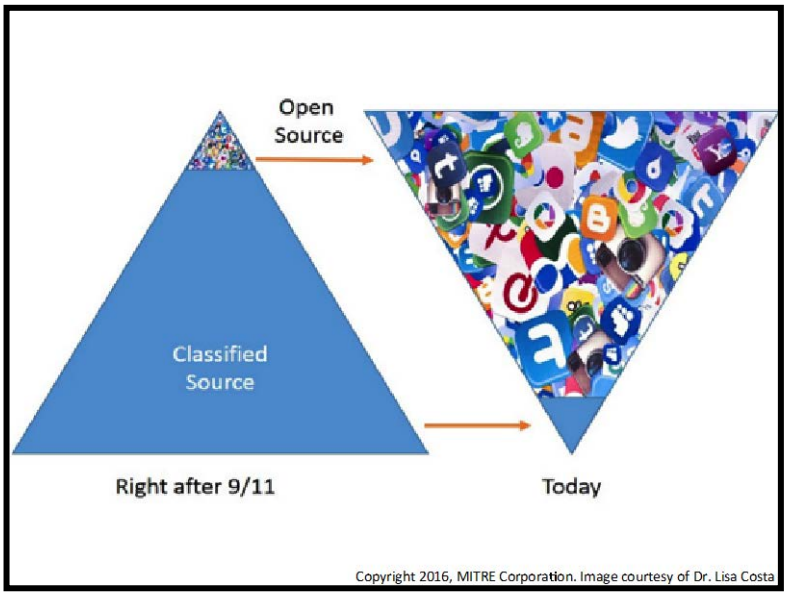
Some intelligence professionals, attempting to gauge and improve ISR effectiveness, look to multi-INT fusion to mitigate this shortfall.[29] Typically this involves the use of several

classified data sources, such as SIGINT and GEOINT, at the onset of information discovery. This INT fusion often does provide greater data fidelity than single-source information, such as SIGINT-only reporting. Unfortunately, this thinking still emphasizes the exclusivity of classified information and promotes the dubious contention that it alone provides a window to truth. There is potential, however, to leverage artificial intelligence and deep-learning algorithms culling openly available digital information to move closer to that goal, especially to the degree that open sources contextualize the environment and contribute socio-cultural nuance. Human-machine collaborative processes using open source information might better provide the analyst the necessary holistic evidence supporting probabilistic conclusions. At a minimum, due to their emerging ubiquity, open sources should be investigated with the same enthusiasm as classified ones.

This is not to say that open sources will, in all cases, provide the crystal ball for intelligence analysis. Nor are they a panacea for all intelligence problems. Each situation requires the requisite examination of its underlying characteristics. But the failure to address open sources' potential merits by reflexively dismissing it for either lack of exotic classification or credibility is, at best, a failure to consider creative solutions. At worst, it signals the IC is unprepared to tackle the emergent complexity of global geopolitical dynamics and risks missing important I&W of future conflict.[30]

The 9/11 attacks provided the impetus for moving open source information into the forefront of the value proposition, in that its ability to significantly augment traditional INTs rapidly became apparent in the counterterrorism mission. However, it was not until the advent of open source Big Data's velocity, variety, and volume characteristics, which became apparent with the explosion of social media, that the potential for greater open source analysis in lieu of an

excessive focus on classified sources became a realistic possibility. One now might consider using open sources as the entry point for the intelligence collection process and using classified data to augment the unclassified source, thus flipping the paradigm upside down. Air Force intelligence personnel must engage in a renewed focus on sources, with special attention given to rapidly expanding openly available data. First, however, it is necessary to understand how much open source data is actually available for this type of analysis. For that, we require a better definition of a term commonly used but seldom understood: Big Data.



Part II: Big Data Defined

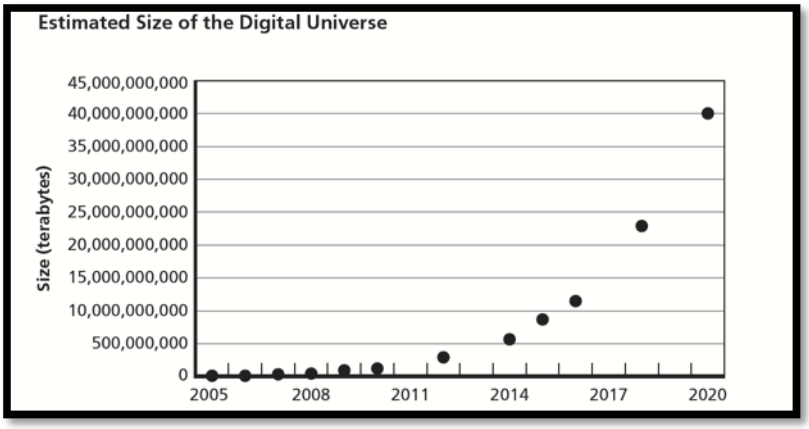
Data Overload Then and Now

In 2010 the U.S. Air Force Deputy Chief of Staff for ISR, Lt Gen (R) David Deptula, lamenting the dilemma facing intelligence professionals, stated that in the very near future they would be "swimming in sensors and drowning in data."[31] Others from fields as disparate as the education, logistics, and environmental policy communities have written similarly.[32, 33, 34] The primary difference between the plethora of data facing intelligence community members as opposed to non-government civilians is that the IC views its problem mainly as it pertains to classified information. General Deptula's own statement, describing "sensors" and the data they provide, underscores that when the military thinks of data it often specifically refers to classified information derived from high-end military platforms. In truth, Air Force intelligence personnel *are* drowning in data. They have only been taught to swim the doggie paddle and are encumbered by classified sensors. Learning creative thinking tools and using machine augmentation while breaking free from the restrictions of classified information can allow analysts to thrive in seas of data.

Since the birth of the internet is relatively recent, it is tempting to think data overload is equally new. However, it is not and it is certainly not new to air forces. As early as World War I, Britain's Royal Air Force collected and distributed more than a million photographs taken from reconnaissance aircraft each month. During a two-month period of the Second World War, the German Luftwaffe took 4,000 photographs each day.[35] Such collection activity likely carried with it an analytical burden analogous to the processing, exploitation, and dissemination (PED) demands the Air Force ISR enterprise faces today. One could argue that the intelligence collection posture of World War I and II combatants was more appropriate. After all, information

regarding the adversary was often much more difficult to acquire then. Moreover, data was not yet digitized, precluding the use of automated tools capable of filtering the signal from the noise available today. Thus, the desire to know more about what is currently unknown predates our own dilemma; but the rush to satisfy that appetite with more collection capability today only serves to compound the problem, creating more data. Luckily, we are finally beginning to look at the analytical challenge of information exploitation as equal to the requirement for collection capability.

The main difference between the volume of ISR collection during the "Industrial Age Intelligence" and what we now face in today's era of "Information Age Intelligence" is scale, also called "volume." Rough calculations estimate that each day we create 10 terabytes of data every .0003 seconds, or 300 microseconds, which is less time than it takes for an iPhone camera to flash (1000 microseconds).[36, 37] In context, that means 90% of the data in the world today has been created in the last two years.[38] A RAND study of cyber vulnerability showed that not only is the digital universe massive, it's getting larger. The study noted, "Estimates of the annual growth of this universe vary, but the increases appear to be exponential."[39] The graphic shows that by 2020 the digital universe will be roughly the size of 40 billion terabytes, a truly unfathomable number. To put it into context, the U.S. Library of Congress is the largest library in the world with 32 million catalogued books and collections that could fill approximately 838



miles of books shelves.[40] Ten terabytes could hold the Library of Congress's entire printed collection.

The term "cyberspace" is commonly used to describe the massive domain where this information exists. Public and private entities alike strive to find the most efficient means to process and analyze the data within it. One part of the Department of Defense's effort to do this is resident within the Third Offset.

The Third Offset: A Technological Requirement

For the past two years, the Office of the Secretary of Defense (OSD) has promoted the Third Offset concept that will allow the U.S. to deter adversary nations from making decisions and executing military deployments contrary to American interests. To meet the intent of providing a deterrent effect, strategic planners and acquisitions personnel must ask in which technologies they should invest. One avenue offering a potentially high return-on-investment is a capability to rapidly provide indications and warning of an adversary's moves toward conflict.

OSD described the Third Offset as the logical follow-on to technological revolutions beginning during the Cold War. In this telling, the First Offset was the introduction of tactical nuclear weapons, which evened the Soviet conventional advantages in the European battle space. The Second Offset, seen during Operation DESERT STORM, brought the introduction of space-based capabilities including the Global Positioning System (GPS), which provided friendly force geolocation and employment of GPS-aided weapons.[41] The Third Offset calls for another technological revolution to provide a leap in capability over potential adversaries. Implicit in OSD's direction is that funding prerogatives will be weighted toward technological advances along five related lines: Learning Systems, Human-machine Collaboration, Human-machine

Combat Teaming, Assisted Human Operations, and Network-enabled Cyber-hardened Autonomous Weapons.

Of the five areas noted in the Third Offset concept, the Air Force ISR enterprise has the greatest interest in deep-learning systems and human-machine collaboration insofar as they assist in providing accurate I&W of impending hostilities. In effect, AF ISR needs to develop, or leverage civilian advances in, learning software that will enable better human-machine collaboration in the Big Data environment to facilitate predictive analysis and prevent surprise. The type of Big Data that offers the greatest utility toward providing future I&W, preventing surprise, and allowing the U.S. to posture deterrent forces will likely be extracted from open sources, not narrowly focused classified sensors. Technological architectures like the Intelligence Community Information Technology Enterprise (IC ITE) will allow lateral information sharing and innovative analyst tools will enable human-machine collaboration. Specific software developments the Air Force is pursuing, such as the Interoperable, Discovery, Exploitation, and Analysis Services (IDEAS) program, should provide the stepping stone to more advanced learning systems. Appendix B at the end of this paper explores both. But before delving into the application of Big Data, first it is necessary to explore the "V's."

Velocity, Variety, and Volume - The V's

Big Data is a term that has been overused to the point that it is often meaningless outside of context. Some simplistic definitions state Big Data is found when more data exists than can populate an Excel spreadsheet; others argue it encompasses generally *all* created and stored information, including text, audio, and visual data. Based on popular information technology trade literature, the most common way of characterizing Big Data is by employing some variation of the 3V's (Velocity, Variety, and Volume) concept.[42] Simply speaking, if you have

concerns about data Velocity, Variety, and Volume, you have a Big Data problem. For a more thorough characterization of this concept, please reference Appendix A.

The 3V's, originally coined in the early 1990s by an Information Technology (IT) analyst at Gartner Inc., denoted the key differences between the large amounts of data created in the Industrial Age compared to that created in the Information Age. As it implied, Velocity denotes the speed at which Big Data moves. Olcutt made an important specification between the velocity one normally associates with internet connectivity and what he called "differential velocity." Differential velocity, in a social media context, refers to the speed at which some content "goes viral" and others do not. Viral videos uploaded to YouTube and microblogs like Twitter can have important contributions to I&W, providing tipping and cueing for further exploration akin to how intelligence personnel sometimes leverage Electrical Intelligence (ELINT) sensors to re-direct imaging sensors. Variety refers to the multitude of information sources an analyst must harness. Typically, intelligence personnel describe these information sources as the "INTs" (SIGINT, ELINT, GEOINT, etc.), weighting them heavily in favor of classified sources over openly available information. In an open source context, Variety includes government reports, academic papers, and commercial imagery among many others. Social media is potentially the most lucrative open source for I&W and worth including at the top of the INT list.[43] Olcutt described the Volume component as the "mind-boggling" amount of information one must sort through.[44] As noted above, the amount of data generated each day increases exponentially. IBM estimates that by 2020, 43 trillion gigabytes of data will be created representing a 300-fold increase from 2005.

Depending on the source, information scientists may include more "V's": Veracity, Variability, Visualization, Value, and even Vulgarity. Appendix A covers more of these

considerations but the question for intelligence personnel remains: will they continue to look to expanding, yet traditional and largely restricted, sources of information,[45] or will they open their aperture to the Big Data potential in open source information? If they choose the latter, then fortunately the IC and DoD have already invested in the technological infrastructure described in Appendix B to facilitate rapid information movement across organizations. However, the technological answer, while necessary, is insufficient by itself for dealing with Big Data. Rather, intelligence analysts require a revision to the ways they think about solving problems. Specifically, they must break from the urge to complete intelligence puzzles that do not exist with more and more classified information that does not always lead to wisdom.

Part III: Open Sources

Flaws in the Technical Solution – Its Not *Only* About Secrets

An unfortunate outcome of the U.S.'s Cold War victory was the misconception that methods previously providing comparative advantages, such as focusing high-end collection systems on scant pieces of information, would be sufficient in the future. Such is the consequence of success: it seldom teaches as good a lesson as failure and it privileges convenience over understanding. Amidst the explosion of openly available digital information, intelligence professionals require awareness that it is a major contributor to understanding intelligence problems and not just a supplement to classified information.

Describing the revolution in information technology computer systems, exemplified by the technological advances described in Appendix B, one intelligence agency leader said, “It’s the foundation for which the community can operate at a faster pace and answer the key intelligence questions that face us today.” That may be true, but this thinking implies that a factual answer is always knowable as long as an individual has access to the right equipment that may deliver the correct information. More important than technical equipment are humans that ask the right questions framed in the correct context.

Speaking to the Council on Foreign Relations, former CIA director General (R) Michael Hayden described intelligence trade as a jigsaw puzzle.[46] The metaphor leads one to believe that all of the pieces are available awaiting assembly. This thinking rewards both the pursuit and creation of more data. Unfortunately, intelligence problems, especially as they pertain to vague indications of impending hostilities, more resemble mysteries. Olcott notes mysteries are difficult, if not impossible, to solve definitively, “no matter how much information is gathered.”[47] At the risk of mixing metaphors, mysteries are more appropriately represented by

abstract paintings, for which the observer (analyst) must use his own subjective judgment to interpret meaning. Trying to answer mysteries usually involves uncertainty, doubt, and cognitive dissonance, which most seek to avoid. But embracing doubt and addressing probabilities are essential because so few intelligence problems lend themselves to easy, certain, factual answers. Philip Tetlock described the allure of certainty, saying that it "satisfies the brain's desire for order because it yields tidy explanations with no loose ends."[48] But Daniel Kahneman warns against the overconfidence certainty can provide: "Declarations of high confidence mainly tell you that an individual has constructed a coherent story in his mind, not necessarily that the story is true."[49] Doubt can sometimes be mitigated, though not eliminated, with more evidence that might even come from classified sources. But pointing to evidence exclusively derived from restricted data while claiming to have found truth is like a blind man describing the colors of a rainbow.

Joseph Nye, describing the challenges facing foreign policy analysts after the Cold War, broke the situation down into a similar dichotomy: mysteries versus secrets. Nye believed that the ratio of the two was increasing in favor of mysteries.[50] He described a mystery as an abstraction that does not lend itself to quick answers or easy analysis. An example might be the likelihood that a foreign leader would pursue a specific course of action in the next year. A more military-oriented question might be whether or not a state will engage in a "hybrid war" campaign employing non-uniformed forces. Secrets, on the other hand, are more defined problems that can be answered via espionage or technical means. They lend themselves to satisfying, seemingly factual, answers and descriptive analysis. Situations requiring expensive sensors for collection such as identifying an adversary's order of battle or the capability of a weapons system abound in military intelligence. However, in the future, as more data becomes

openly available, proper employment of human-machine collaboration may yield insights formerly reserved to classified sensors alone.

While puzzles requiring the acquisition of secret "pieces" do persist, leveraging open source information is increasingly able to help us better understand mysteries *and* answer specific, defined problems. Social media usage by "first responders" in the form of the civilian populace may be able to address questions traditionally answered by military-operated collection assets. Open sources can point to breakthroughs in a nation's weapons R&D timeline, a task formerly the exclusive province of espionage or technical sensors. To this point, Anthony Olcott notes that intelligence puzzles asking questions like "What are the range and speed capabilities of the latest generation Chinese surface-to-air missile (SAM)?" can be addressed through the lens of open sources. He illustrates the commercial, public sector, use of open sources, relating what Leonard Fuld calls the cardinal rule of intelligence: "Wherever money is exchanged, so is information." All of the components required for successful SAM operations conceivably leave paper trails exploitable through financial and commercial analysis. This point was further supported in an interview the author conducted with Dr. Robert Norton of Auburn University. Dr. Norton described the potential of monitoring foreign weapon manufacturing through an adversary country's research and development timeline. Dr. Norton noted that foreign universities emphasize the need to publish in technical journals as much as American higher education centers do. Information concerning technical developments can often be observed slowly building and then rapidly disappearing, perhaps marking that a country has reached the appropriate phase of research to begin transitioning a capability to the operational test and evaluation phases.

However, as Nye predicted, intelligence problems of the future will likely more resemble mysteries than secrets. With that likely eventuality intelligence analysts will need to employ more creative and critical thinking and use a more diverse assortment of information than before. This will entail greater mental workload for analysts used to the collect-process-analyze model traditionally centered on the classified collection. The IC's knee-jerk inclination to accept the answer offered by classified data satisfies what Daniel Kahneman calls our System 1 response, a mode of thinking in which the mind operates "automatically and quickly, with little or no effort."[51] Kahneman contrasts this mode with the concentration required of System 2, which "allocates attention to the effortful mental activities that demand it."[52] The uncertainty created by nebulous mysteries that often don't lend themselves to prompt answers, and the creative thinking required to solve them, is System 2 territory.

Josh Kerbel described the undue emphasis on classified information as a barrier to creative thinking. He argues that classified collection carries the “need to know” restriction, which “fosters compartmentalized – reductionist – view of the issues at hand.”[53] He also points to the Cold War era “when (the IC) had a relative monopoly on good information” which “continues to cause analysts to confuse exclusivity of information with relevance to decision-makers.”[54] This also points to a key distinction regarding information availability then and now. During the Cold War, prized information was often technical and ephemeral, mainly consisting of communications and/or electrical emissions. Intelligence professionals often refer to this data as a “detectable signature” of the collection target. Such detectable signatures, fleeting during the Cold War, have exploded in the Information Age.

Terms like the “Internet of Things” (IoT) refer to networked, interoperable connectivity. One way the IoT manifests itself today is in the growth of wearable technology. It often provides

precision geolocation of an individual and connects one's previously private details to the open architecture of the internet. Moreover, this information does not have to be secretly seized by a high altitude sensor or through clandestine espionage. In fact, individuals *willingly* make their data available for observation. This is why, as Treverton noted, in the Information Age, "collecting information is less of a problem, and verifying it is more of one."[55] The open source environment provides detectable signatures of the adversary undreamed of prior to the advent of the Information Age. By comparison, the lucrative collection environment offered by the IoT makes sporadic Cold War-era collection of machine emissions quaint. Nevertheless, the IC and conventional military's nearly exclusive focus on Industrial Age collection targets, and the corresponding construct of privileged access it fosters, persists. The net result is that the exclusivity of secret intelligence becomes the basis for analysis to the limitation, or even exclusion of, creative thinking.*

The military intelligence professionals and the leadership they work for are caught in a double-blind classification paradox. Access to classified information carries with it the currency of prestige. Intelligence analysts drowning in data assume that they are swimming in a sea of wisdom because of their exclusive access to restricted information. Meanwhile those in leadership positions with limited access to classified information, or without the time to study it, assume that those with access really know the truth and are good stewards of the data. Both are often engaged in willful self-deception.

---

* As Lt Col Adam Stone identified in a 2016 Air War College study, the Air Force does not have the luxury of tapping into a wealth of critical thinking (CT) capability to begin with. Pointing to the Air Force Future Operating Concept's desire for the identification of critical thinkers and metrics to track critical thinking skills, Stone executed a quantitative CT research project. He tested a sample of PME students in-residence at Air Command & Staff College (ACSC), the School for Advanced Air and Space Studies (SAASS), and Air War College (AWC). His (statistically significant) results indicated, "AF officers attending ACSC and AWC were below average in CT skills when compared with individuals at the same academic level." (Stone, Adam "Critical Thinking Skills of US Air Force Senior and Intermediate Developmental Education Students.")

The influx of open source data, including rapidly growing social media platforms, will only become more vital sources of information in the future. Kerbel continues, "[The IC] must get over its now illusory belief that its value-added comes mostly from information to which it alone has access – secrets."[56] Several open source venue social media companies are billion-dollar-a-year companies, to wit: Instagram. Founded in 2010, it is arguably the fastest growing social media channel reaching 300 million users in 2016.[57] Alec Ross noted, "Today there are roughly 16 billion internet connected devices. Four years from now that number will grow to 40 billion internet-connected devices."[58] Open source data, rather than closely guarded classified information, is ***the*** Big Data challenge facing intelligence professionals.

Despite this openly available source of information, the term OSINT remains a lesser INT in the realm of intelligence disciplines. This is largely due to a classification fixation, historical bias, and joint doctrine that deems it merely a supplement to classified collection. Openly available information is often treated as less trustworthy than covertly gathered information, but this confuses the source (openly available information) with the product (open source intelligence). Libor Benes quoted a retired CIA officer offering an unfortunate yet prevailing, view: "By definition, intelligence is clandestinely acquired information – stolen, to put it bluntly."[59] Such is the thinking of an outdated mindset. Offering the contrasting view, Treverton stated, "Intelligence now has… vast amounts of information… not a scarcity of information that mainly comes from satellites or spies and is therefore regarded as accurate."[60]

Openly available information is not only a valuable supplement; it is redefining strategic I&W and should be used as the basis of future intelligence analysis. In essence, open sources should not augment secret information, but the reverse.

Indications & Warning – The Open Source Risks

It is important to remember that the internet is an equal opportunity provider of open source data; U.S. forces are as susceptible to its capabilities as any other internet user. This fact became real to many in the DoD after the May 2011 Bin Laden strike when a Pakistan-based IT consultant inadvertently live-tweeted the raid.[61] In the midst of one of the most OPSEC-intensive operations the U.S. military has ever conducted, the following tweet appeared for literally anyone in the internet-connected world to see:

Again in 2011, Operation ODYSSEY DAWN brought with it another education in vulnerability for U.S. forces. Owing to the Automatic Dependent Surveillance-Broadcast (ADS-B)[62] avionics upgrade, anyone with "a general knowledge of (air traffic control) procedures… off the shelf electronics, and the internet"[63] could track multiple coalition force aviation assets. In essence, any internet user can now build his/her own air surveillance picture – a capability formerly reserved for those with access to classified data. In a recent AFIT paper, Major Donald McCallie wrote, "This example demonstrates the efforts of only one individual; consider what a motivated, funded adversary could achieve. ADS-B implementation will only make it cheaper, easier and faster to accomplish what once took many assets to achieve."[64]

These two examples highlight an important shift in information availability. Prior to the ubiquity of the internet and hand-held mobile devices, information mainly flowed from the top of a state's hierarchy downward. In authoritarian regimes, the state controlled the levers of information dissemination and thus defined "reality" for those under its control. In that environment, it was worthwhile to focus on official state pronouncements and propaganda as they might allow insight into an adversary's mindset at the national level. In *Open Source Intelligence in a Networked World,* Olcott described the transition from tightly controlled information dissemination of the Cold War to the influx of information available today: "The balance of power in the communication relationship (shifted) from the producer to the audience." Communication is no longer a top-down push process. Increasingly, individuals are able to pull the information they want from the places that best fit their specific proclivities. This makes it tremendously difficult for authoritarian regimes to manage dissent, but at the same time provides a lucrative avenue for intelligence analysis. Olcott continues, "The ability of almost anyone to send information, and of people to choose from whom they wish to receive it, has played havoc with the ways (governments) have controlled 'reality' within their boundaries." Unlike anytime during the Cold War, information dissemination at the individual level provides any one person the agency to participate in the information environment and define reality for himself.

Mining Open Source Big Data & Machine Augmentation

Special Operations Forces (SOF) are making tremendous strides with respect to leveraging social media Big Data in support of operations.[65] Lessons learned from unconventional operations show strong Big Data approaches are built around data-driven processes that use Information Technology (IT) to its fullest potential. Namely, IT should be used to develop repeatable processes to replace some human-based approaches, such as reading

raw data. Also, IT should help analysts understand what the data say without having to read each source. The MITRE Corporation, in support of special operations, uses Big Data private industry tools for social media exploitation "to understand how populations feel about their conditions, leaders, and political groups."[66] This focus on "populations," as opposed to higher leadership echelons, highlights the balance of power shift that Olcott described. Moreover, it points to the utility of allocating classified collection to supplement unclassified information sources. The diffusion of communication power from formerly rigid hierarchies to the masses illustrates how a focus on the latter may be the best avenue for intelligence analysis seeking to provide I&W and prevent surprise.

Dr. Lisa Costa, an analyst with extensive experience supporting Special Operations Command (SOCOM), argued that the IC must understand and access the emerging world that is becoming rapidly social media-based. She notes that "over time, the amount of OSINT of value has increased to a volume and velocity that classified sources cannot compare to from a corroboration, access, and economic perspective."[67] To that end, Dr. Costa recommended that SOCOM specifically adopt a 95/5 rule whereby 95% of the information analyzed is open source and only 5% classified.[68] This paper does not argue that such an extreme ratio is necessary for conventional operations, but conventional intelligence analysts would do well to heed SOCOM's lessons. In the future, the lines between unconventional and conventional operations will continue to blur.

I&W – The Open Source Opportunity in "Hybrid War"

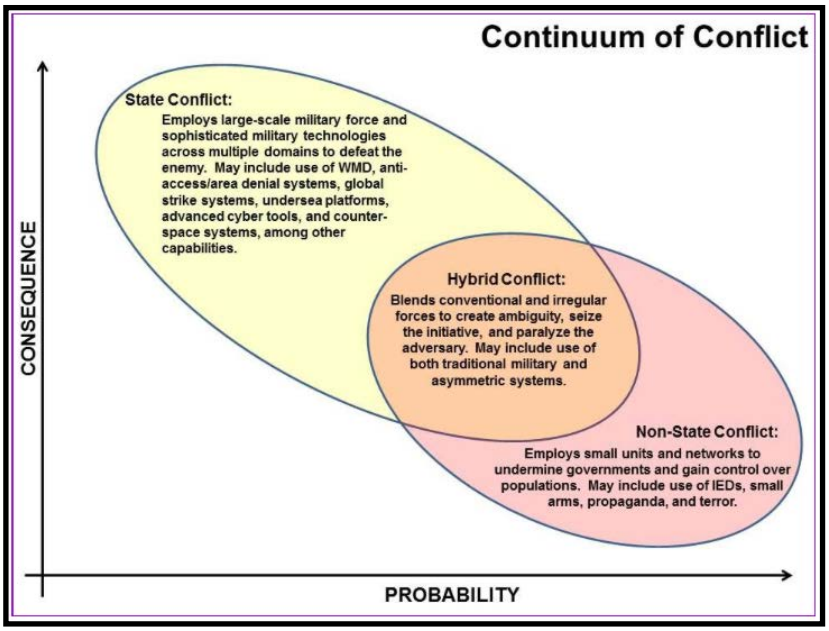
In the summer of 2014 "pro-Russian separatists" began appearing in eastern Ukraine. Moscow repeatedly denied that its regular forces were operating on Ukrainian soil, but social media truth gave lie to the Russian government's insistence. Young soldiers posted "selfies" to

Instagram that, presumably unbeknownst to them, contained metadata that geolocated their position within Ukraine's borders. This type of cueing indication in and of itself does not constitute an "end product" – but it could be used to direct traditional ISR collection assets to verify the open source tip.

Later in the summer of the 2014 Russia-Ukraine conflict, a civilian aircraft carrying nearly 300 passengers was shot down by an advanced Russian-built surface-to-air missile (SAM). Multiple open source reports after the event led to the overwhelming conclusion this SAM system was a modern weapon not typically associated with insurgent forces. Normally, this would be a situation tailor-made for the U.S.'s Cold War ISR architecture to detect and geolocate the conventional weapon. But initial accounts rapidly flowed in from open sources to include pictures uploaded to Twitter and Instagram as well as numerous YouTube videos.[69] Despite the crushing weight of evidence to the contrary, Russian defense outlets improbably assigned blame to Ukrainian forces.[70] Russian President Vladimir Putin, seizing the opportunity to win the propaganda war, crisply stated as much, declaring the situation could have been avoided "if Kiev had not resumed its military campaign against pro-Russian separatists."[71] Such rhetoric is the hallmark of what is becoming known as "hybrid war." These types of conventional and unconventional incidents mixed with intense public relations campaigns will be the U.S. military's most likely, and most dangerous, scenarios for conflict into the future.

The former Chairman of the Joint Chiefs of Staff, CJCS GEN (R) Martin Dempsey, offered (also outlined in the 2015 National Military Strategy, see graphic) an appropriate definition of hybrid war: "State and non-state actors working together toward shared objectives, employing a wide range of weapons such as we have witnessed in eastern Ukraine."[72] He continues, "Hybrid conflicts serve to increase ambiguity, complicate decision-making, and slow the coordination of effective responses. Due to these advantages to the aggressor, it is likely that this form of conflict will persist well into the future."[73]

Some have argued the concept of hybrid war in the Ukraine is simply a continuation of conventional techniques and procedures.[74] Whatever the definition, what we are seeing is the most likely scenario for future conflict because it allows the adversary to do as Sun Tzu recommended: capitalize on the adversary's weaknesses while maximizing its own strengths. The U.S. military possesses overwhelming conventional might. But by engaging in disinformation and employing non-official military forces, the adversary can keep the conflict below the threshold where the U.S. might use its conventional advantage. As some have noted, such hybrid war strategies could "cripple a state before that state even realizes the conflict had begun," and yet it manages to "slip under NATO's threshold of perception and reaction."[75]

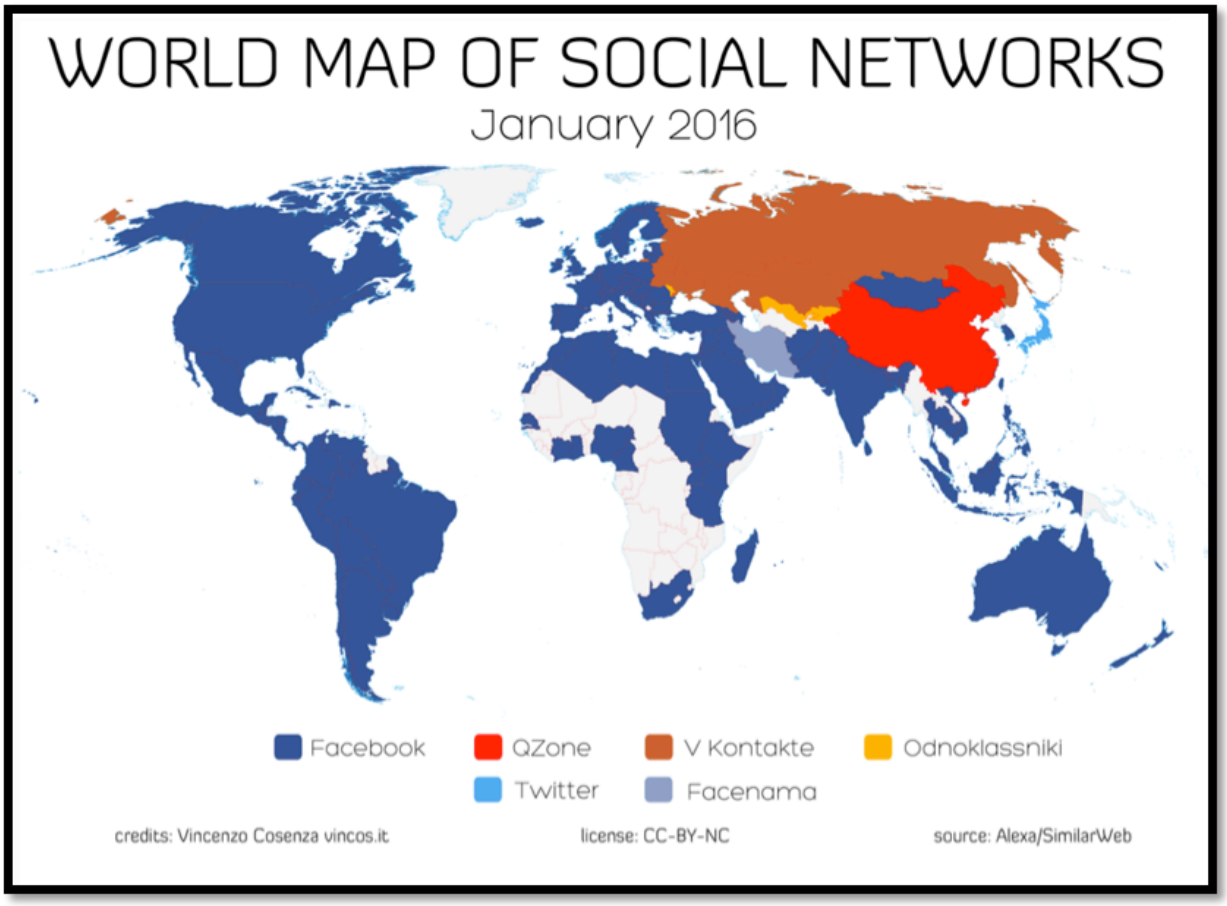


Hybrid war is at the same time the most dangerous scenario because potential aggressors, such as Russia and China, possess significant military resources to escalate conflict. Both countries maintain nuclear weapons and intercontinental delivery systems.[76] Unlike other sub-state or quasi-state organizations that may engage in hybrid war (Hezbollah, ISIL, etc.), state-driven hybrid war that escalates to a nuclear exchange poses existential threats to humanity.

Hybrid wars of the future will not only be confined to the U.S.'s adversary in the European plain. Michael Pillsbury, a defense expert with 40 years' experience, described at length China's possible employment of a military doctrine called "unrestricted warfare."[77] This doctrine in many ways is analogous to hybrid warfare, especially with respect to finding asymmetries against a conventionally stronger foe. Pillsbury described China's aggressive maritime claims in the East and South China seas.[78] One possible dangerous course of action in this region might entail the use of "civilian" Chinese fishing boats executing what for all intents and purposes is a military operation to solidify its assertion of territoriality. Such activities are not without precedent. A Council on Foreign Relations White Paper states, "In June 2011, Vietnam accused a Chinese fishing boat of cutting cables from an oil exploration vessel inside its [economic exclusion zone]."[79] Small-scale "fishing incidents" may become the source of increasing naval tensions. They provide China plausible deniability while remaining under the threshold for spurring greater U.S. military involvement.

As Steven Pifer noted in 2015 testimony before the U.S. Senate, irregular forces presaging larger conventional movements may be the example of future hybrid war encounters.[80] General Philip Breedlove, the EUCOM commander, also articulated these concerns, insisting NATO must be prepared to respond to "special forces without sovereign insignia who cross borders to create unrest" and ultimately destabilize countries.[81, 82] However, traditional I&W

techniques and ISR systems employed by the IC have historically focused on the deployment of large armed forces. They do so at the risk of missing earlier indications that might forestall conflict. As stated, OSD's intention in pursuing the Third Offset concept is to deter potential adversaries from action. The AF ISR enterprise's goal toward that end should be providing the timeliest I&W of impending conflict to avoid decision paralysis as the U.S. confronts entities falsely claiming noncombatant status. A sub-goal should be to provide decision makers evidence to counter aggressor propaganda. The best tool for these missions in the future will likely not be a traditional collection platform originally designed to count Soviet tanks; it will be open source-derived information. Cold War era TTPs are not conducive to identifying "little green men," innocuous fishing vessels, nor the funding, arms, and leadership supporting them. Therefore, a change to the way the AF ISR Enterprise conducts operations is warranted. That begins with a focus on open source information augmented by secret-seeking sensors capable of adding detail resulting in open source intelligence.



There is an important role for human-machine collaboration in this new open source-focused environment, specifically with respect to artificial intelligence (AI). While social media outlets like Facebook, Instagram, and Twitter are popularly used worldwide, some countries use other social media outlets more predominantly. The graphic above shows that the social networking sites V Kontakte and QZone are the most popular outlets in Russia and China, respectively. Analysts must be cognizant of that fact and adept at deciphering not only foreign languages but cultural nuances of the society in question. Automatic machine translation tools are rapidly improving and can help with both. In May 2014, Microsoft presented a computer program capable of translating spoken words in real time.[83] Describing the application of "deep learning" to machine translation, Maryam Najafabadi, *et al.* relate how Google's "word2vec" tool can quickly learn complex relationships between hundreds of millions of words.[84] Using

what are called “word vectors” allows the machine translator to distinguish nuance and context rather than literal translation. AI translation tools directed at social media outlets could provide a wealth of insight into what Olcott refers to as lower-level authority structures. With more communicative power now in the hands of lower-level individuals, the biggest obstacle is “simply deciding where best to pay attention in the ever-more cacophonous bazaar of would-be message senders now competing for the attention of the world, and of each other.”[85] Machine augmentation will not only allow us to hear what these individuals are saying, but understand what they mean.

Part IV: Conclusion and Recommendations

The Way Forward for the AF ISR Enterprise

Major changes are required in the way intelligence professionals think about problems. A cultural mindset change is warranted that values freely available information as much as, if not more so than, restricted data. For the Air Force, change will begin at entry-level education and training venues, namely officer and enlisted initial skills training courses at Goodfellow AFB. Due to differences in generational familiarity with technology, this will likely be the easiest step. "Digital natives," the next generation of intelligence professionals that have grown up with ubiquitous technology and social media outlets, will likely find it easier to break from legacy mindsets. However, the lure of the classified source will still be seductive. Intelligence training must support the next generation's inclination to reach for the open source.

Additionally, future Airmen will require training in the tools available at that time and encouragement to pursue their own innovative ideas to best collect and analyze open source material. The app-store-like capabilities within IC ITE and JIE (see Appendix B) will facilitate this. But as Robert Folker wrote, technological solutions must be combined with Airmen trained to use analytical techniques.[86] Specific analytic training should include problem restatement, causal flow diagramming, weighted rankings, devil's advocacy, and many other techniques as described by Morgan Jones in *The Thinkers Toolkit: 14 Powerful Techniques for Problem Solving*. Empirical evidence concludes, "Exploitation of a structured methodology will improve qualitative intelligence analysis."[87] Follow-on programs like the Advanced Analysis course teach these techniques. However, advanced analytical courses are not compulsory and only attended by a small fraction of the overall pool of Air Force intelligence personnel. The importance of teaching analytical techniques to Airmen for use with the avalanche of data cannot be over-

emphasized. First, these techniques allow the analyst to "show their work," making their analyses transparent to others. Second, they teach language precision, forcing the analyst to frame the problem correctly to ensure it is answerable and not open to interpretation. Lastly, they can prevent military analysts from falling into the System 1 trap that Kahneman described. The natural human inclination to grab onto the first plausible explanation is a key challenge for anyone, but especially for intelligence professionals confronted with the time constraints of military operations. This problem leads to the recurring spiral of the double-blind classification paradox noted above. Those in leadership positions will often seek data compatible with the beliefs they already hold. Normally, in military operations, this means a desire for classified information over less glamorous open sources.[88] Kahneman defines this as confirmation bias and one way to break free of it is to use skills inherent in applying appropriate analytical techniques. Investing in the development of critical thinking skills is an expressed requirement outlined in the Air Force Future Operating Concept. With these skills and knowledge Airmen will be able to better respond to Combatant Commanders and provide more decision-quality material, rather than wasting time and effort on mundane production quotas endlessly seeking puzzle pieces.

It is perfectly reasonable to be skeptical about the utility of employing slow, methodical techniques to improve analytical rigor. But if the Air Force is serious about developing critical thinking skills, the right answer is not to dismiss these techniques out of hand but to experiment in accordance with proven scientific methods. In a small-scale examination, the AF ISR Enterprise could give the training on different analytical techniques to a randomly selected group of analysts, but not another, and compare the results. Tetlock notes, "The intelligence community's forecasters have never been systematically assessed"[89] to determine the accuracy of their analytic predictions. To this author's knowledge, neither have the U.S. Air Force's. The

entire test would be relatively inexpensive compared to the cost of flying and maintaining ISR platforms. Looking at the results of this experiment could provide valuable, low-cost information that might better enable future planning and budgetary decisions.

Critical thinking skills bolstered by proven analytic techniques will make human-machine pairing a more lethally effective combination. But as human-machine collaborative tools become more ubiquitous in the future, training courses must not teach the Airmen analyst to rely solely on machine-based solutions and accept the answer as truth. In *The Shallows: What the Internet is Doing to Our Brains*, Nicholas Carr warns against the neurological effects of our increasing reliance on technology. He writes that such reliance limits creative thought, "preventing us from achieving the intellectual depth that leads to wisdom."[90] Instructors at Goodfellow AFB should expect and require the machine augmentation they employ to help students visualize the data. Proper visualization tools, like certain components of the IDEAS software (reference Appendix B), may help ensure the students understand the machine's process and why it presents the conclusions it offers to maximize human-machine collaboration. The AF ISR Enterprise must use the talents and open source inclinations our Airmen of the future bring to formal training, but also teach them how best to employ them.

This begs a question, though, about "legacy" intelligence analysts. Is there an analytical role for pre-9/11 personnel in the future information environment? Based on anecdotal evidence, individuals born prior to the Information Age will likely be less welcoming of open source material and more disposed to favor traditional sources of collection. Jervis wrote, "Those who are most involved in carrying out politics guided by the old image will be the least able to innovate."[91] But rather than endure the slow movement of time while the next generation ascends to leadership positions, forward thinking leaders must break from the classification fixation now.

They must realize the relevance of open sources to guide collection, not the other way around. Additionally, individual analysts must *want* to contribute. But how?

The Intelligence Advanced Projects Activity (IARPA) is an ODNI-sponsored program that challenges participants across the IC to engage in forecasting competitions. A spinoff program called the Good Judgment Project involves any willing participant both inside and out of the DoD. The first IARPA tournament began in 2011 and explored the potential of crowd-sourced intelligence. Participants made predictions about real-world events, which were then judged by the precision of their forecast. Perhaps the most interesting outcome of the Good Judgment Project was that individuals with access to restricted information had no advantage over those without. In fact, the opposite was true, possibly due to the cultural bias towards classified information that may have prevented those individuals from forming more holistic predictions. In a Washington Post Op-Ed detailing the competition's results, David Ignatius specified that individuals *without* access to classified information "performed 30% better than the average for the intelligence community analysts who could read intercepts and other secret data."[92] He continues, "The NSA obviously operates on the theory that more data are better…but this mad dash for signals lacks the essential quality of sound judgment."[93]

Benjamin Franklin said, "Tell me, and I forget. Teach me, and I remember. Involve me, and I learn."[94] Just as the Air Force stresses physical training (PT) culminating in regular biannual or annual tests, so should the AF ISR enterprise champion regular "cognitive PT" tournaments. Results from multiple Good Judgment Project competitions revealed, "Prediction accuracy is possible when people participate in a setup that rewards only accuracy -- and not the novelty of the explanation, or loyalty to the party line."[95] In other words, competitions like these foster both creative and critical thinking while honing skills on an individual level. Furthermore,

competitions may lend themselves to developing and asking questions that can be answered, measured, and scored. This may carry over to analysts supporting operational commanders. When an analyst faces an open-ended question that cannot be easily answered, he will be more likely to analyze the question, breaking it down into its component parts, and developing a series of smaller, measureable questions that will lead to answering the original question asked, but with more precision and accountability. Competitive events are not new for the Air Force. For decades fighter pilots have trained against rival squadrons during Turkey Shoot events. Winners receive accolades and the recognition of their peers. The Air Force ISR Enterprise needs an ISR Turkey Shoot, challenging participants to form their own conclusions, thereby granting agency to the individual and allowing motivated Airmen to best demonstrate their analytic prowess.

The dilemma in building these analytical competitions is in framing the question correctly. Scope the question out too far ("How will conflict in Region 'X' play out?) and the answer's accuracy becomes difficult to judge. But these big picture questions, often framed in sweeping generalities, are what often matter most to policy makers and military leaders. Ask a question too narrowly ("Will Country 'X' launch a ballistic missile in the next six months?") and the answer may be trivial. The key, then, is two-fold: questions must be relevant to the Air Force, and of sufficient scope to offer a meaningful answer. Tetlock offers a "deductivist" recommendation to develop questions meeting both of these criteria. He notes that the "big question" is often comprised of smaller ones capable of providing worthwhile insights: "And if we ask *many* small-but-pertinent questions, we can close in on an answer for the big question." [96] Will Country X employ non-uniformed personnel in a new area of operations? Will it increase its territorial claims? Tetlock notes the questions are cumulative and "the more yeses, the likelier

the answer to the big question" (i.e. How will conflict in Region X play out?) is "This is going to end badly."[97]

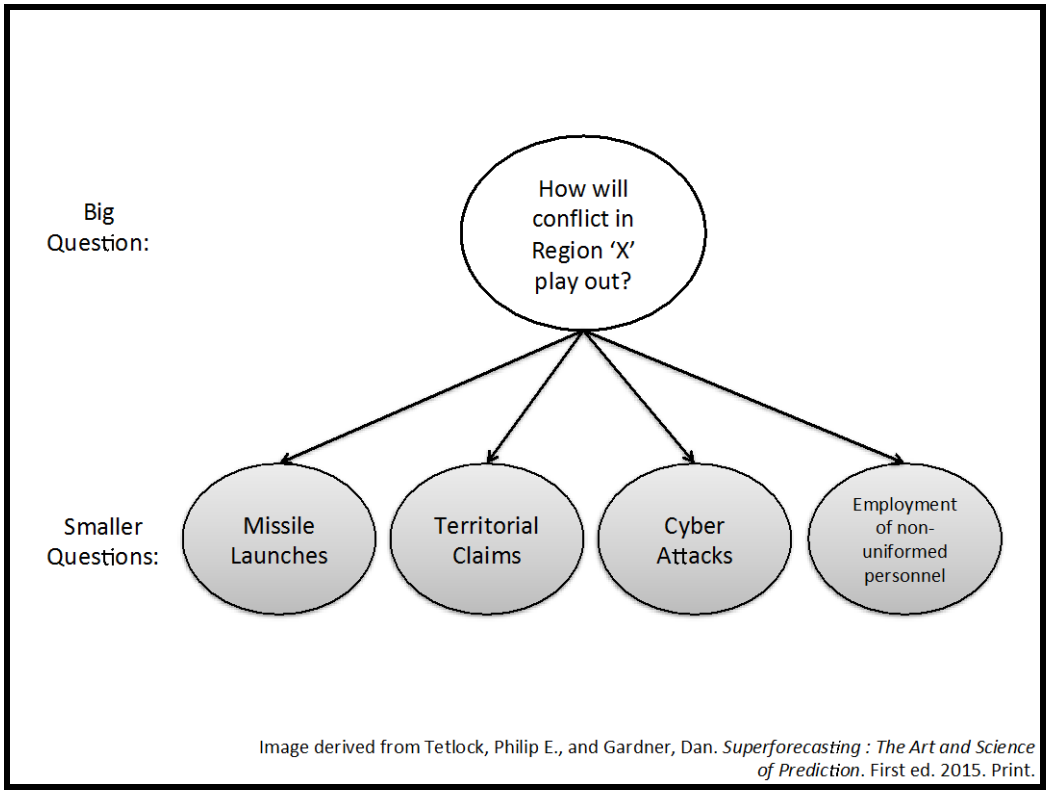


Image derived from Tetlock, Philip E., and Gardner, Dan. *Superforecasting : The Art and Science of Prediction*. First ed. 2015. Print.

For relevance, it is important to frame the question based on Air Force interests. But framing the question correctly and answering it are radically different. The deductive approach presents problems worth addressing. It implies a formal logic model whereby one reasons based on known premises, or premises presumed to be true (i.e., more missile launches mean an adversary more intent on conflict escalation), and then follows them to their logical conclusion.[98] The conclusions are thus more certain, or could be presented as such. As described previously, this manner of thinking best suits the closed problem sets and intelligence puzzles of the Cold War where the cause and effects were more closely linked.[99] Thus, this is a dangerous pathway for intelligence analysis today and requires more creative solutions. Analysts should also use inductive approaches to the problems. Inductive conclusions are often listed as probable or

plausible but stop short of identifying definitive causality.[100] Perhaps the aggressor state is employing non-uniformed forces from a sense of weakness, not strength. Such is the gray world presented by hybrid warfare environments. To assist the analytical process, analysts must go beyond traditional notions of causality and traditional classified information sources with their seemingly helpful puzzle pieces.

To avoid bias in favor of exclusive access to restricted information, the tournament should focus on information readily available in the open source domain. Individual participants or even teams of all ranks and experience could enter with the incentive that their successful performance would result in Air Force-level recognition. Currently, the Air Force ISR enterprise manages the AF ISR Awards Program (AFISRAP), which honors exceptional contributions to the field. Headquarters Air Force (HAF) should regard winners of this cognitive competition as the highest level of intelligence excellence. Further, HAF/A2 should work with other interested Air Force Specialties, challenging them to participate in this program on a non-interference basis. Doing so could encourage anyone to test his or her analytic skills, opening up a larger audience for information crowdsourcing. HAF/A2 could also approach Joint partners to expand the program to the rest of the DoD and reap more of the same benefits. Competition between individuals and units could spur motivation and breed further intelligence excellence. And, based on the results of the Good Judgment Project, one might expect open source disbelievers to become converts. At a minimum, participants will learn that classified sources matter less then the rigor one applies to analysis. These cognitive PT tournaments should form the basis for future Air Force-centric prediction markets asking answerable questions relevant to Air Force decision makers. Tracking these results over time would allow the Air Staff to determine

whether some individuals and organizations consistently predict accurate results, thereby showing whether the entire endeavor is bearing fruit.

Lastly, another interim method the AF ISR enterprise can employ to eliminate its cultural bias against unclassified sources is to follow the footsteps blazed by the integration of kinetic and non-kinetic effects. As the Air Force came to realize the threat to air operations from contested, degraded, and operationally-limited (CDO) environments, planners began to welcome discussions of non-kinetic employment. Operational planning discussions now routinely feature full-spectrum (kinetic and non-kinetic) solutions for mission execution because the problem was framed in terms of the CDO environment. As a component of the cognitive competition described above, a similar approach could work for intelligence purposes using I&W. Intelligence professionals could compete to find the best open source solutions to the I&W problem in the open information environment. Framing the problem as one of information naturally leads to information solutions, of which OSINT is paramount due to the rapidly proliferating digital universe.

Final Points – An Opportunity for Success

The IC and DoD investment in technology is important, but without a mindset change that recognizes the potential value of open source Big Data and corresponding investments in analytical training and creative thinking, it is a wasted expenditure, as if an individual bought a Ferrari without taking driving lessons. The individual might get where he is going, but it would be bumpy, inefficient, and ultimately a poor investment decision. It is tempting to assume that a machine capable of crunching mass quantities of data will automatically produce a result sufficient to render the human irrelevant. The hype regarding data science reinforces this

seductive possibility. The science fiction writer Arthur C. Clarke said as much in his "third law": "Any sufficiently advanced technology is indistinguishable from magic."[101]

If Air Force intelligence professionals succeed in the future it will not be due to magical machines, but through the difficult, patient process of creatively analyzing problems and articulating viable solutions. The data feeding those solutions will increasingly be found in readily accessible, yet traditionally stigmatized, open sources. As LTG (R) Michael Flynn wrote while serving as the senior intelligence officer in Afghanistan, "The intelligence community's standard mode of operation is emphatic about secrecy but regrettably less concerned about mission effectiveness."[102] Individuals must overcome the classification fixation and focus on the information that best leads to mission success.

Machines can assist the analytical process but they are not a substitute for it. In *Superforecasting*, Tetlock presents a pessimistic outlook for machine-only analysis and an inversely optimistic appraisal for the future of human cognition. He states, "Machines may get better at 'mimicking human meaning' and thereby better at predicting human behavior,"[103] but he argued there is a significant difference between mimicking meaning and deciphering the meaning's original intent. He concludes, "That's a space human judgment will always occupy."[104] To that end we must invest in the human mind in the form of analytic training. Targeted investments directed toward improving creative thinking and smartly leveraging open source Big Data will ultimately assist leadership decision-making and allow the Air Force a strong comparative advantage in the future.

Appendix A: The V's of Big Data

Velocity refers to the speed with which the incoming data arrives and multiplies. However, it also speaks to the speed required to link data sets to other data as quickly as it arrives, connecting the dots. This is truly a difficult challenge in today's digital environment where individuals send 200 million emails every minute, tweet 500 million times per day, and Google queries run to 1.2 trillion searches per year.[105] Military and civilian examples of Velocity are similar; both realms must contend with increasing data flows without a tremendous amount of distinction. But from a military standpoint Velocity also implies a need to both receive and process the information so that it is rendered "decision quality"; that is, actionable. The military requires automated systems that can move at the speed of the data to quickly recognize situations that potentially demand a lethal response. This is particularly true if intelligence personnel are to provide timely, predictive I&W.

Variety denotes the breadth of data sources and types. Some of it is "structured" which allows processors to more easily ingest it, and some of it is "unstructured," which does not. The degree to which the data is structured heavily influences how rapidly it can be retrieved for analysis. Variety in the military context refers to the multiple sources of information flowing from platforms and sensors, some owned and controlled by the DoD and some not. From the Air Force perspective these might be air-breathing platforms (U-2, RC-135, fighter aircraft, etc.) or space-based satellite assets. Increasingly, however, the variety of sources also includes open source material. The variety of open material today, social media most notably, is flipping the traditional sources of data upside down. The challenge for ISR professionals in the future is to consider open sources at least to the same degree as classified sources, especially in light of their staggering potential for intelligence analysis.

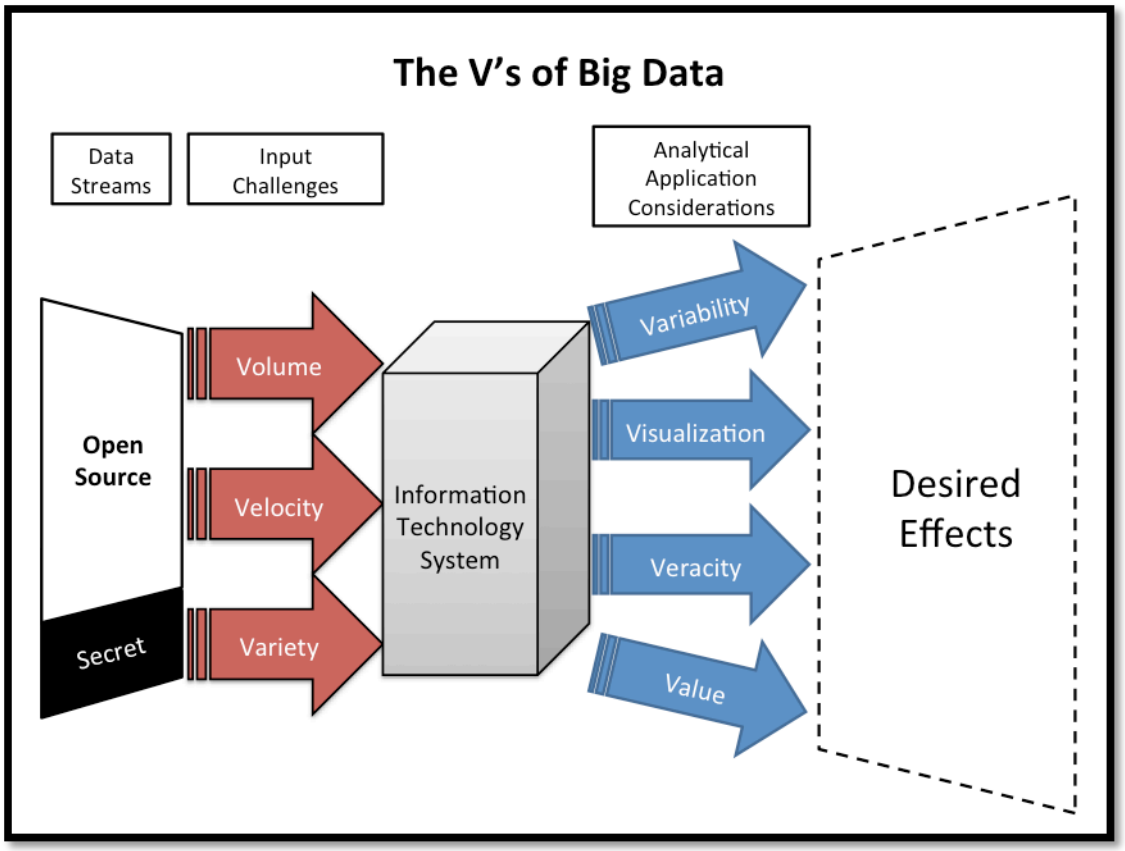
Volume gets to the heart of the Big Data problem. The major difference between the Big Data of the past and today is the degree to which we are able to store massive quantities of information for little cost. For instance, it costs Google, arguably the world leader in Big Data analytics, $0.01 per gigabyte of storage.[106] Nate Silver argues one could trace the storage problem back to the printing press. Gutenberg's invention meant the accumulation of knowledge was no longer dependent on the ability of an individual to memorize texts.[107] For our purposes it is more apt to point to the development of the microchip in the 1960s, which paved the way for digital storage. Put simply, cheap Big Storage enables Big Data. One observes the military's Volume dilemma most acutely with respect to the full-motion video explosion. One consequence of the land-centric OIF and OEF wars of the 2010s was the insatiable demand for live video feeds. Storing terabytes of that data is a challenge for the USAF but solutions may exist in digital technology upgrades this paper addresses in Appendix B.

Four More V's – the *Application* of Big Data

Some IT publications have expanded the 3V's concept to include four additional V's: Veracity, Variability, Visualization, and Value. The difference between the original 3V's and the four add-ons is that the former better represent Big Data in terms of scale while the latter four represent the problems inherent in *applying* Big Data toward analytical solutions. Nevertheless, these additional considerations are important to explore in order to proceed toward a better understanding of the challenges of Big Data and Big Data analytics.

Veracity is the degree to which the data is trustworthy. It bears directly on the usefulness of the Big Data analytic end product if the product is built on false information. This will be of increasing importance in the future as malicious cyber activity threatens to deliberately manipulate data.[108]

The Variability of the data is potentially the greatest challenge technology must overcome before true consistent analysis is possible. Not to be confused with Variety, Variability speaks to rapidly changing contexts within which the data appears.[109] For instance, for learning software to be useful in a Big Data environment it must be able to delineate meaning and decipher nuance in the data it receives, something at which humans typically excel. As Gary Marcus writes in the New Yorker, Big Data works well in systems that are consistent over time with established properties, small variation, and minor complexity. Unfortunately, these criteria



are not well suited to the world of battlefield intelligence where the fog of war predominates. Pointing to the limitations of Big Data with respect to its predictive power, he warns, "Big Data is a powerful tool for inferring correlations, not a magic wand for inferring causality."[110] Marcus points to the difference between the IBM's "Watson" supercomputer's capability to answer Jeopardy questions, which are essentially a data-retrieval function, and the challenges of chess strategy. The latter require a contextual understanding of each position's best move dependent on complex relationships with the other pieces. Human-machine collaboration appears to be the answer to the "Variability" problem,* as Deputy Secretary of Defense Robert Work noted in 2015 remarks at the Reagan National Defense Forum.[111,] He pointed to a 2005 chess tournament where amateur players using their own computers were able to beat their supercomputer opponents through a mix of human ingenuity and machine augmentation.

Building on the Third Offset concept of human-machine teaming, Visualization is the method by which Big Data makes itself accessible to the human analyst. A 2013 journal article by Ashleigh Faith highlighted the importance of visually depicting data. She noted, "Visualization tools create an 'atmosphere of opportunity' wherein data relationships can be transformed into knowledge."[112] Essentially, visualization tools allow the analyst to connect seemingly unrelated data points. Visualization also allows the machine to "show its work" to the analyst so that the latter can explain how the machine arrived at its conclusions. Thus, visualization is the essential element of human-machine collaboration.

* In a very recent, and startling, update with respect to artificial intelligence (another component of the Third Offset) on March 12, 2016 a Google supercomputer beat the world's top 'Go' player three games in a row. This is significant because the game Go requires a level of intuition and creativeness unlike chess. The Los Angeles Times put it succinctly, "(it) brings to a close the era of board games as benchmarks in computing." (http://www.latimes.com/world/asia/la-fg-korea-alphago-20160312-story.html)

The final V according to IT professionals is Value. This component, somewhat related to Veracity, represents the difference between useful and non-useful information. Simply put, the data must be relevant and useful for Big Data analytics or it results in wasted time and effort. It is important to note, though, that the data in and of itself is irrelevant without the (largely human) analytical component that must work in concert with the technological developments that bring the data to the analyst.

Appendix B: Emerging Technological Architectures

While the IC is not faced with the same degree of denied information that it was during the Cold War, there exists a need and use for a traditional ISR capability to find answers to some "secrets." Countries still maintain tight operational security over some of their most sensitive programs and many are a national security threat to the United States. In the future, these instances will likely decrease, as more data becomes openly available. But when those occasions arise, analysts can bring to bear the overwhelming bulk of the IC bureaucracy, which was designed to gather and analyze secrets. The problem then is how can the bureaucracy, with its agencies historically closed to one another, share information effectively?

One Big Data problem for the ISR community echoes the stovepiping dilemma just discussed. In fact, stovepiping compounds the Big Data problem. Individual IC agencies have problems with Volume, Variety, Velocity, etc. and on top of that they are poorly integrated, deceasing the ability to share data. Each member of the IC generates enormous amounts of raw (sometimes called "unfinished") intelligence as well as finished products. But organizational barriers prevent merging this data into a cloud-like architecture for multispectrum data analysis and integration. To correct this issue, both the IC and DoD are engaged in a technological revolution represented by three separate but related programs. They are the Intelligence Community Information Technology Enterprise (IC ITE), the Joint Information Environment (JIE) and the Defense Intelligence Information Enterprise (DI2E).

Intelligence Community Information Technology Enterprise

IC ITE is a program led by the Office of the Director of National Intelligence (ODNI) and focused on the intelligence community specifically. It will "integrate classified data repositories and workflow across the IC."[113] Relative to the other initiatives this paper describes,

it is a smaller endeavor. It is designed specifically to break down the barriers between intelligence agencies to improve efficiency. IC ITE will do this by using a single IC IT infrastructure rather than separate IT architectures for each individual agency. The end goal of IC ITE is to "establish a powerful platform to deliver more innovative and secure technology to desktops at all levels across the (IC)."[114] To enable this, ODNI created smaller, manageable tasks for some of the agencies to complete for the benefit of the rest of the IC.

The first sub-task was to build the common desktop. NGA and DIA are implementing the "IC Common Desktop" that will "provide a uniform interface and (enable) analysts at any agency to communicate and exchange information."[115] In a two-phase approach NGA and DIA will deploy these new systems within their agencies before next delivering them to the rest of the IC. The common desktop will allow each agency a common visualization tool. To address the core big three (Variety, Volume, and Velocity) the IC looks to the cloud.

A second critical IC ITE component is the cloud architecture, which the CIA and NSA are building through their Commercial Cloud Services (C2S) and GovCloud, respectively. In an interesting link between private and public enterprises, the CIA is working directly with Amazon, who developed the C2S program on behalf of the Agency.[116] In fact, both organizations are looking to commercial cloud-based solutions from Google, Microsoft, and Amazon to address a multitude of concerns including network security and reliability.[117] Perhaps most importantly, the IC is looking at all of this with an eye toward a "pay for play" concept rather than a flat fee regardless of use which will put the onus on the private sector to create value-added capabilities. This type of partnership will be crucial in the future as the IC leverages commercial entities' comparative IT advantage.

Joint Information Environment

Driven by a 2012 document called "Capstone for Joint Operations: Joint Force 2020," the Joint Information Environment will attempt to create the same type of technological efficiencies IC ITE seeks. While IC ITE focuses exclusively on the intelligence side of IT, JIE has a larger scope "centered on IT upgrades at the secret and unclassified levels for the entire DoD." To that end JIE addresses operations, maintenance, and support requirements. Further, it plans to create an enterprise solution – a single joint platform – for all the members of the DoD. To do that JIE must also address the storage issue.

The foundations of the JIE are processing and storage centers called Joint Regional Security Stacks (JRSS). According to DISA the JRSS comprise the physical stacks for the DoD's cloud architecture and will "enable big data analytics, allowing DoD components to intake large sets of data to the cloud and provide the platforms for processing the data."[118] JRSS's also address cyber security concerns as the JRSS removes the requirement for each base or post to conduct localized network security.

The JRSS also includes a new routers called Multiprotocol Label Switching (MLS) equipment. They will upgrade the DoD's bandwidth and decrease the instances of stalled or lost connectivity. Notably, the MLS equipment is designed to cope with high-volume incidents when connectivity is crucial to successful operations.

Lastly, DoD is looking to private industry to facilitate cloud services similar to IC ITE's approach. JIE's "milCloud" is a DISA-managed infrastructure that combines commercial off-the-shelf components and government developed technology. Unlike the JRSS, which provides information storage and retrieval, milCloud will allow DoD an app store-like feature where users can place orders for applications they desire for their particular operational focus.[119]

DI2E (Defense Intelligence Information Environment)

Where IC ITE and JIE provide solutions for the IC and DoD, respectively, DI2E provides the link between them. DI2E provides the "common framework of standards, processes, technologies, and reference implementations" so that the IC and DoD can build and share apps and information across their sub-communities. For example, DI2E allows a COCOM logistics officer easier access to information in an IC database that may be of interest to his or her mission. Furthermore, through another app-store program called DI2E Developers Environment, users can build and test new applications collaboratively for free.

Making all of this integrate seamlessly is the DI2E council, a joint interagency governing body with representatives from across the DoD and IC. Security concerns form a major hurdle to bridge the IC ITE and JIE worlds. As most users of classified information know, JWICS connectivity is less available further down the operational chain and, inversely, many IC components spend less time on secret and unclassified systems. The DI2E Council is addressing this system security mismatch in order to provide mission effectiveness to the warfighter.[120]

IC ITE, JIE, and DI2E will all help to alleviate the problems caused by institutional stovepiping, but outside of the Volume consideration they do not directly address Big Data's V's. Creative individuals and organizations within the IC and DoD need to build the software that will fit into the overall technological architecture. To that end the U.S. Air Force has developed a program of its own.

The Air Force Has IDEAS

Proposed technological solutions must contend with the Volume, Variety, and Velocity of Big Data information. Machine augmentation via the National Air & Space Intelligence Center's (NASIC) Interoperable, Discover, Exploitation, and Analysis Services (IDEAS) program may address the outstanding Big Data V's not addressed by IC ITE, JIE, and DI2E.

The IDEAS program is a suite of interoperable system-of-systems software built to ingest, condition, transform, and make discoverable large volumes of information trapped within the world of Big Data.[121] It operates within an IC-ITE compatible architecture allowing it to exchange information across the IC. To overcome "Variety" challenges, IDEAS was built to process multimodal (i.e., text, images, video, and audio) data. According to the program's developer, Mr. James Homer, IDEAS's greatest challenge is to "transform and condition" the data.[122] Conditioning refers to the ability to process unstructured data most notably images, video and audio which "typically lack rich exploitable meta-data."[123] IDEAS's ability to transform and condition the data allows for follow-on exploitation and collaboration dependent on the needs of the user.* This automation ultimately reduces the need for manual intervention. IDEAS addresses Variability and Visualization concurrently through human-machine collaboration. Analysts can immerse themselves in the data, viewing spatial and temporal clusters of related information based on user queries that allow the data to speak for itself. Through visualization, humans contribute nuance and context, deciding for themselves if the results are valuable for further analysis.

Programs like IDEAS are not, however, an end unto themselves but rather a means to provide information to decision makers during critical phases of conflict. The rapidly progressing capabilities of near-peer adversaries makes the pre-hostilities phase of war increasingly important. Joint Publication 3.0 describes "Phase 0" as the period of conflict when "[Joint Force Commanders] are able to assist in determining the shape and character of potential future operations before committing forces."[124] To win the Phase 0 fight in future conflicts, the

* According to a NASIC Whitepaper IDEAS has, "demonstrated a capacity to inject and condition digital and hardcopy English, Russian, and Chinese open source literature, S&T journals, conference papers/notes, videos, images, classified message traffic, finished intelligence, and email."

Air Force will have to contribute to adequately shaping the environment. Given the forward deployed forces present during the Cold War, reaction time was more on the U.S.'s side than now. The Velocity and Variety of information must be analyzed to quickly select the "signals" from the terabytes of open source "noise." Therefore, the AF ISR enterprise requires human-machine collaboration and learning systems like IDEAS to provide early warning to Joint Force Commanders. Fortunately, given the technological advances outlined above, the infrastructure will soon exist to share vast amounts of both restricted and open source information across communities. NASIC's IDEAS program can put tools in the hands of intelligence professionals to provide timely and decisive decision advantage for JFC's.

The IC realizes that Cold War stovepipes designed to push I&W of surprise attack upward to decision making elites at the expense of sharing the information laterally has outlived its usefulness. The IT revolution underway spanning the IC to DoD will helpfully break the barriers between these organizations. Users will shortly be able to work in concert and access the breadth of U.S. ISR Big Data with the help of IC ITE, JIE, and the governing DI2E protocols. Innovations like the IDEAS program will augment human capabilities allowing users to see and exploit the data.

Technology will fix the Cold War technological problem, but Industrial Age warfare mindsets remain. However valuable the technological fix, cultural problems persist. Current ISR operations favor the production of arguably low-utility materials over knowledge and classified sources over openly available information. Thus, the technological architecture is necessary, but insufficient, for future mission success.

Notes

1. I wish to thank Col Jeffrey Donnithorne, Dr. Jon Kimminau, Dr. Lisa Costa, Dr. Robert Norton, Mr. Josh Kerbel, Mr. James Homer, Lt Col Robert Folker, and Majors Kyle Bressette and Seth Gilpin for their thoughtful comments and suggestions. All errors found therein are my own.
2. Adam Lowther and John Farrell. "From the Air." *Air & Space Power Journal* 26, no. 4 (2012): 61-102.
3. According to the Office of the Director of National Intelligence, the U.S. Intelligence Community budget (in terms of appropriated dollars) for the most recent year of reporting (2014) was $67.9 billion divided between the National Intelligence Program and the Military Intelligence Program. See http://www.dni.gov/index.php/intelligence-community/ic-policies-reports/ic-policies-2?highlight=WyJidWRnZXQiXQ==
4. Treverton, Gregory F. *Reshaping National Intelligence for an Age of Information*. RAND Studies in Policy Analysis. Cambridge ; New York: Cambridge University Press, 2003.
5. Kerbel, Josh. "The U.S. Intelligence Community's Creativity Challenge," The National Interest, 13 October 2014, http://nationalinterest.org/feature/the-us-intelligence-communitys-creativity-challenge-11451
6. Ibid.
7. Baer, R. (2002). *See no evil : The true story of a ground soldier in the CIA's war on terrorism* (1st ed.). New York: Crown. Also see Hoffman, David E. 52.1 (2010): 220. Print.
8. Treverton, G. (2003). *Reshaping national intelligence for an age of information* (RAND studies in policy analysis). Cambridge ; New York: Cambridge University Press.
9. Olcott, Anthony. *Open Source Intelligence in a Networked World*. London and New York,Continuum, 2012. Print.
10. Treverton. p 7.
11. Throughout the Cold War the national-level imagery agency changed names several times from the National Photographic Interpretation Center (NPIC) to the National Intelligence & Mapping Agency (NIMA) to, ultimately, NGA.
12. Treverton, p 8.
13. Ibid.
14. Josh Kerbel, "The U.S. Intelligence Community's Creativity Challenge," The National Interest, 13 October 2014, http://nationalinterest.org/feature/the-us-intelligence-communitys-creativity-challenge-11451
15. Joint Publication 2-0, *Joint Intelligence*, 22 October 2013, B-7.
16. Olcott, Anthony, *Open Source Intelligence in a Networked World* (London and New York: Continuum, 2012) Kindle Edition, chap 4.
17. Ibid.
18. Joint Publication 2-0, *Joint Intelligence*, 22 October 2013, B-7.
19. Lingel, Sherrill Lee, Rand Corporation, Project Air Force, Mu'assasat Rānd, Rand, RĖND, SShA, and Rėnd. *Methodology for Improving the Planning, Execution, and Assessment of Intelligence, Surveillance, and Reconnaissance Operations*. Santa Monica, CA: RAND Corporation, 2007. http://www.rand.org/content/dam/rand/pubs/technical_reports/2008/RAND_TR459.pdf
20. Lt Col David Vernal, (Air War College Student), interview by the author, 15 November 2015.
21. Ibid.
22. It is worth noting that that Joint Pub 2-0 does not include these considerations regarding veracity for any other intelligence discipline other than OSINT. This again reinforces a belief that only the other, non-OSINT 'INTs', are capable of providing truth.
23. Olcott, Anthony, *Open Source Intelligence in a Networked World* (London and New York: Continuum, 2012) Kindle Edition, chap 4.
24. Allison, Graham T., and Zelikow, Philip. *Essence of Decision: Explaining the Cuban Missile Crisis*. 2nd ed. New York: Longman, 1999.
25. Jervis, Robert, Harvard University. Center for International Affairs, and Cfia. *Perception and Misperception in International Politics*. Princeton, N.J.: Princeton University Press, 1976.
26. As taught from 1998-99 at the Joint Military Intelligence College course ANA630, "Intelligence Analysis: Continuity and Change," and derived from Ronald D. Garst, "Chapter 4: Fundamentals of Intelligence Analysis;" reprint in *Intelligence Analysis: ANA630* vol. 1 (Washington: Joint Military Intelligence College, nd.), 18-28.

27. The difference between descriptive analysis describing who, what, and where questions and predictive analysis is that the latter does not easily lend itself to statements of fact, thereby inducing greater cognitive stress. Addressing predictive "what will happen" questions forces the analyst to assume more risk by making subjective judgments amid uncertainty. Moreover, these open-ended questions must be answered with a range of possibilities, often tied to equally subjective probabilities.
28. Kahneman, Daniel. *Thinking, Fast and Slow*. 1st Pbk. ed. New York: Farrar, Straus and Giroux, 2013. Print.
29. Reference endnote 46 regarding this point.
30. Josh Kerbel, "The U.S. Intelligence Community's Creativity Challenge," The National Interest, 13 October 2014, http://nationalinterest.org/feature/the-us-intelligence-communitys-creativity-challenge-11451
31. Magnuson, Stew. "Military 'Swimming In Sensors and Drowning in Data'." *National Defense* [Arlington] 01 Jan. 2010: 36-38. Web.
32. Briggs, Beverley. "Drowning in Data." *The Times Educational Supplement* [London] 27 Sept. 2013: 14. Web.
33. Lancaster, Roy, Mac Talbert, and Rebecca Kirk. "'Drowning in Data, Starving for Information'." *United States Naval Institute. Proceedings* 140.2 (2014): 78-79. Web.
34. Ottinger, Gwen. "Drowning in Data." *Issues in Science and Technology* 27.3 (2011): 71-82. Web.
35. Treverton, Gregory F. *Reshaping National Intelligence for an Age of Information*. Cambridge ; New York: Cambridge UP, 2003. Print. RAND Studies in Policy Analysis.
36. *Wikipedia*, "Microsecond", accessed 10 March 2016, https://en.wikipedia.org/wiki/Microsecond
37. This figure was taken from an IBM estimate that 2.5 quintillion are created each day which equates to $2.5x10^9$ Terabytes per day which also equates to 104,166,666 Terabytes per hour. Accessed 10 March 2016, http://www-01.ibm.com/software/data/bigdata/what-is-big-data.html
38. Ibid.
39. Porche, Isaac R.. Emerging Cyber Threats and Implications. Santa Monica, CA: RAND Corporation, 2016. http://www.rand.org/pubs/testimonies/CT453.html.
40. Matt Raymond, "How 'Big' Is the Library of Congress?" United States Library of Congress, 11 February 2009, https://blogs.loc.gov/loc/2009/02/how-big-is-the-library-of-congress/
41. Cheryl Pellerin, "Work: Human-Machine Teaming Represents Defense Technology Future," U.S. Department of Defense News, 8 November 2015, http://www.defense.gov/News-Article-View/Article/628154/work-human-machine-teaming-represents-defense-technology-future
42. Some have argued a fourth V (Veracity) should be included to properly address Big Data problems. However, the author concurs with Doug Laney that veracity is inversely related to data "bigness." See http://blogs.gartner.com/doug-laney/deja-vvvue-others-claiming-gartners-volume-velocity-variety-construct-for-big-data/ , also http://www.forbes.com/sites/gartnergroup/2013/03/27/gartners-big-data-definition-consists-of-three-parts-not-to-be-confused-with-three-vs/#2b7d523c3bf6, and additional V's at https://datafloq.com/read/3vs-sufficient-describe-big-data/166
43. Some organizations have taken to calling social media intelligence SOCMINT. This paper does not advocate the concept of creating another 'INT' but this shows a prevailing recognition of social media's usefulness to some intelligence communities.
44. Olcott, Anthony. *Open Source Intelligence in a Networked World*. London and New York,Continuum, 2012. Print.
45. The term "Activity Based Intelligence" (ABI) is gaining notoriety in the IC as a possible way to deal with Big Data challenges and is worth addressing here. IT professionals created the original 3V's, and the application of the succeeding 4V's, attempting to describe the situation they faced with respect to commercial business interests. Lt Col Chandler Atwood discusses, in a paper describing ISR applications of Big Data, the AF ISR enterprise acutely faces similar data overload challenges. He notes, "Even today in Afghanistan where ISR forces have been redundantly layered for years, the creation of a timely, coherent picture gained from multisource intelligence data is a rarity." (Atwood, Chandler. "Activity-Based Intelligence: Revolutionizing Military Intelligence Analysis." *Joint Force Quarterly : JFQ*, no. 77 (2015): 24-29,32-33.) He goes on to describe that the V's of Big Data "requires a significantly different way of handling the task(s) that traditional intelligence methodologies cannot support." (Ibid) He illustrated an example showing that the Intelligence Community had foreknowledge of an impending chemical weapons attack in Syria but "failed to process and integrate the information in time to portend such an attack." (Ibid) He points to the 'stovepiping' of information, whereby intelligence collection is maintained in separate channels for analysis (NSA for SIGINT, NGA for GEOINT, etc.), which yields poor interoperability. To deal with the Big Data problem and mitigate gaps in single-source collection Lt Col Atwood advocates for ABI. The ABI methodology focuses on fusing classified sources of information early in the information discovery process to provide an analyst

with a full picture of information from which he begins to form conclusions. While ISR fusion is necessary and ABI is a worthwhile methodology, both still overly invest an undue amount of attention on classified sources as the key driver in making sense of the collected information.

46. Gen Michael Hayden (remarks by Central Intelligence Agency Director at the Council on Foreign Relations, Washington D.C., 7 September 2007), https://www.cia.gov/news-information/speeches-testimony/2007/general-haydens-remarks-at-the-council-on-foreign-relations.html
47. Olcott, Anthony. *Open Source Intelligence in a Networked World*. London and New York,Continuum, 2012. Print.
48. Tetlock, Philip E., and Gardner, Dan. *Superforecasting : The Art and Science of Prediction*. First ed. 2015. Print.
49. Kahneman, Daniel. *Thinking, Fast and Slow*. 1st Pbk. ed. New York: Farrar, Straus and Giroux, 2013. Print.
50. Nye, Joseph S. "Peering into the Future." *Foreign Affairs* 73.4 (1994): 82-93. Web.
51. Kahneman, Daniel. *Thinking, Fast and Slow*. 1st Pbk. ed. New York: Farrar, Straus and Giroux, 2013. Print.
52. Ibid.
53. Kerbel, Josh http://warontherocks.com/2016/01/the-u-s-intelligence-community-wants-disruptive-change-as-long-as-its-not-disruptive/
54. Ibid.
55. Treverton. p 9.
56. Ibid.
57. Harris, Matthew. "Marketing with Instagram, the Fastest Growing Social Platform!" The Medium Well, 19 February 2016, http://mediumwell.com/marketing-instagram/
58. Alec Ross, "Industries of the Future" (lecture, Carnegie Council podcast, 10 March 2016).
59. Benes, Libor. "OSINT, New Technologies, Education: Expanding Opportunities and Threats. A New Paradigm." *Journal of Strategic Security* 6, no. 3 Suppl. (2013): 22-37.
60. Treverton. p 6.
61. Griffith, Erin. "12 Tweets That Changed the World," Fortune Magazine, 8 March 2016, http://fortune.com/2016/03/08/tweets-changed-world/
62. ADS-B provides a continuous broadcast of an aircraft's position, identity, and velocity over unencrypted datalinks or air traffic management.
63. McCallie, Donald L., and Air Force Institute of Technology . Graduate School of Engineering Management. *Exploring Potential ADS-B Vulnerabilities in the FAA's NextGen Air Transportation System* (2011). Web.
64. Ibid.
65. Costa, Lisa. "Large Scale Data Analytics for Operational Decision Making." Powerpoint presentation.
66. Ibid.
67. Ibid.
68. Ibid.
69. Foreign Policy Staff, "What We Know So Far About the Passenger Jet Shot Down in Ukraine," Foreignpolicy.com, 17 July 2014, http://foreignpolicy.com/2014/07/17/what-we-know-so-far-about-the-passenger-jet-shot-down-in-ukraine/
70. Ibid.
71. Ibid.
72. Marcus Weisgerber, "Dempsey's Final Instruction to the Pentagon: Prepare for a Long War." *Defense One* (blog), 1 July 2015, http://www.defenseone.com/management/2015/07/dempseys-final-instruction-pentagon-prepare-long-war/116761/
73. Ibid.
74. Kofman, Michael. "Russian Hybrid Warfare and Other Dark Arts," Warontherocks.com, 11 March 2016, http://warontherocks.com/2016/03/russian-hybrid-warfare-and-other-dark-arts/
75. Schadlow, Nadia. "The Problem With Hybrid Warfare," Warontherocks.com, 2 April 2015, http://warontherocks.com/2015/04/the-problem-with-hybrid-warfare/
76. Butler, Declan. "Obama's Nuclear-weapons-free Vision." *Nature* 458.7239 (2009): 684-5. Web.
77. Pillsbury, Michael. *The Hundred-year Marathon : China's Secret Strategy to Replace America as the Global Superpower*. First ed. 2015. Print.
78. Ibid.
79. Xu, Benia, "South China Sea Tensions," Council on Foreign Relations, 14 May 2014, http://www.cfr.org/china/south-china-sea-tensions/p29790

80. Pifer, Steven. "Russian Aggression Against Ukraine, and the West's Policy Response." *Hampton Roads International Security Quarterly* (2015): 23. Web.
81. NATO flexes its muscle memory; the future of NATO. (2014, Aug 30). *The Economist, 412*, 55-56. Retrieved from http://aufric.idm.oclc.org/login?url=http://search.proquest.com.aufric.idm.oclc.org/docview/1558845880?accountid=4332
82. Gonzales, D., & Harting, S. (2014, 04). Exposing russia's covert actions. *U.S.News & World Report,* , 1. Retrieved from http://search.proquest.com.aufric.idm.oclc.org/docview/1527433272?accountid=4332
83. "Rise of the Machines; Artificial Intelligence." *The Economist* 415.8937 (2015): 18-21. Web.
84. Najafabadi, Maryam, M. Villanustre, Flavio Khoshgoftaar, Taghi Seliya, Naeem Wald, and Randall Muharemagic. "Deep Learning Applications and Challenges in Big Data Analytics." *Journal of Big Data* 2.1 (2015): 1-21. Web.
85. Olcott, Anthony. *Open Source Intelligence in a Networked World*. London and New York,Continuum, 2012. Print.
86. Folker, Robert D., and Joint Military Intelligence College. *Intelligence Analysis in Theater Joint Intelligence Centers : An Experiment in Applying Structured Methods*. Washington, D.C.: Joint Military Intelligence College, 2000. Print. Occasional Paper (Joint Military Intelligence College (U.S.)) ; No. 7.
87. Folker, and Joint Military Intelligence Coll Washington DC Center FOR Strategic Intelligence Research. "Intelligence Analysis in Theater Joint Intelligence Centers: An Experiment in Applying Structured Methods." (2000). Web.
88. Kahneman, Daniel. *Thinking, Fast and Slow*. 1st Pbk. ed. New York: Farrar, Straus and Giroux, 2013. Print.
89. Tetlock, Philip E., and Gardner, Dan. *Superforecasting : The Art and Science of Prediction*. First ed. 2015. Print.
90. Carr, Nicholas G. *The Shallows : What the Internet Is Doing to Our Brains*. 1st ed. New York: W.W. Norton, 2010. Print.
91. Jervis, Robert, Harvard University. Center for International Affairs, and Cfia. *Perception and Misperception in International Politics*. Princeton, N.J.: Princeton University Press, 1976.
92. David Ignatius, "More Chatter Than Needed," Washington Post, 1 November 2013.
93. Ibid.
94. Karen Angelo, "Internships and Co-op Opportunities Expand," University of Massachusetts-Lowell, 15 October 2012, https://www.uml.edu/News/stories/2011-12/ServiceLearning.aspx
95. Chen, Angela. "Philip Tetlock's Tomorrows." The Chronicle of Higher Education [Washington] 05 Oct. 2015: The Chronicle of Higher Education, Oct 5, 2015. Web.
96. Tetlock, Philip E., and Gardner, Dan. *Superforecasting : The Art and Science of Prediction*. First ed. 2015. Print.
97. Ibid.
98. Levin-Rozalis, Miri. "Using Abductive Research Logic: "The Logic of Discovery", to Construct a Rigorous Explanation of Amorphous Evaluation Findings." *Journal of MultiDisciplinary Evaluation* 6.13 (2010): 1-14. Web.
99. Josh Kerbel, "The U.S. Intelligence Community's Creativity Challenge," The National Interest, 13 October 2014, http://nationalinterest.org/feature/the-us-intelligence-communitys-creativity-challenge-11451
100. Levin-Rozalis, Miri. "Using Abductive Research Logic: "The Logic of Discovery", to Construct a Rigorous Explanation of Amorphous Evaluation Findings." *Journal of MultiDisciplinary Evaluation* 6.13 (2010): 1-14. Web.
101. Ibid.
102. Flynn, Michael, Matthew Pottinger, and Paul Batchelor. "Fixing Intel in Afghanistan." *Marine Corps Gazette* 94.4 (2010): 62-67. Web.
103. Tetlock, Philip E., and Gardner, Dan. *Superforecasting : The Art and Science of Prediction*. First ed. 2015. Print.
104. Ibid.
105. "Google Search Statistics," accessed 10 February 2016, http://www.internetlivestats.com/google-search-statistics/
106. Futures Laboratory Roundtable Discussion, 19 Feb 2016
107. Silver, Nate. *The Signal and the Noise : Why so Many Predictions Fail--but Some Don't*. New York: Penguin, 2012. Print..

108. DNI Clapper warned as much in testimony before the House Intelligence Committee as did the NSA director Adm Rogers during a Senate Intelligence Committee hearing. http://thehill.com/policy/cybersecurity/254977-officials-worried-hackers-will-change-your-data-not-steal-it
109. "Variability is often confused with variety. Say you have bakery that sells 10 different breads. That is variety. Now imagine you go to that bakery three days in a row and every day you buy the same type of bread but each day it tastes and smells different. That is variability." - https://datafloq.com/read/3vs-sufficient-describe-big-data/166
110. Marcus, Gary. "Steamrolled by Big Data," The New Yorker, 29 March 2013, http://www.newyorker.com/tech/elements/steamrolled-by-big-data
111. Another building block, human-machine collaboration, was in the news in 1997 when IBM supercomputer Deep Blue beat chess grandmaster Garry Kasparov -- the first defeat of a current world chess champion to a computer under tournament conditions. Then in 2005, Work said, "two amateur chess players using three personal computers won $20,000 in a chess tournament against a field of supercomputers and grandmasters." http://www.defense.gov/News-Article-View/Article/628154/work-human-machine-teaming-represents-defense-technology-future
112. Faith, A., M.I.L.S. (2013). Information intelligence: A blueprint for data visualization. *Information Outlook (Online), 17*(3), 15-18. Retrieved from http://search.proquest.com.aufric.idm.oclc.org/docview/1372161540?accountid=4332
113. Quinn, Kristin. "A Global Intelligence Enterprise," Trajectorymagazine.com, accessed 15 February 2016, http://trajectorymagazine.com/government/item/2116-a-global-intelligence-enterprise.html
114. Office of the Director of National Intelligence. "IC ITE Fact Sheet." (http://www.dni.gov/files/documents/IC%20ITE%20Fact%20Sheet.pdf)
115. Quinn, Kristin. "A Global Intelligence Enterprise," Trajectorymagazine.com, accessed 15 February 2016, http://trajectorymagazine.com/government/item/2116-a-global-intelligence-enterprise.html
116. Konkel, Frank. "The Details About the CIA's Deal With Amazon," The Atlantic magazine, 17 July 2014, http://www.theatlantic.com/technology/archive/2014/07/the-details-about-the-cias-deal-with-amazon/374632/
117. National Security Agency, "An Overview of Cloud Computing," https://www.nsa.gov/research/_files/publications/cloud_computing_overview.pdf
118. Defense Information Systems Agency, "Joint Regional Security Stacks," accessed 2 March 2016, http://www.disa.mil/initiatives/jrss
119. Defense Information Systems Agency, "milCloud," accessed 2 March 2016, http://www.afcea.org/events/jie/14/documents/MILCLOUD_MARTIN--FINAL.pdf
120. Quinn, Kristin. "A Global Intelligence Enterprise," Trajectorymagazine.com, accessed 15 February 2016, http://trajectorymagazine.com/government/item/2116-a-global-intelligence-enterprise.html
121. IDEAS Whitepaper
122. From an interview with the author and Mr. James Homer. One could delineate between transforming and conditioning data in the following way:

"Typically data transformation is use to describe converting statistical data, or in the process the process of converting data from one format (e.g. a database file, XML document, or Excel sheet) to another. I see these transformations as those which can be accomplished via manipulation of the format. In our case I see us transforming the objects into a different state e.g. image of text to text, audio to transcript, one language to another (e.g. Chinese to English), etc.

Data conditioning is the application of data management and optimization techniques to enable the "system" to convey and orchestrate the data objects across the backbone infrastructure and between adjacent services respectively.

Example: a PDF image of Chinese text. First I transform the image via the OCR process from image to text. In our case we then take the Chinese text and transform it into English via machine translation. Now we have both a Chinese and English text versions which are now transformed from the text format into a now normalized XML schema. The data is now considered conditioned because it has been normalized so it can be managed and optimized while transiting the IDEAS infrastructure and between adjacent services under."
123. Ibid.
124. Joint Publication 3-0, *Operations*, 11 August 2011, xx