SYM-AM-18-081

PROCEEDINGS
OF THE
FIFTEENTH ANNUAL ACQUISITION RESEARCH SYMPOSIUM

THURSDAY SESSIONS
VOLUME II

Acquisition Research:
Creating Synergy for Informed Change

May 9–10, 2018

March 30, 2018

Prepared for the Naval Postgraduate School, Monterey, CA 93943.

ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL

Toward Cognitive Supremacy via Quantitative Augmentation

Col Clark Quinn, USAF—is the Chief of the Strategic Planning Integration Division, Headquarters U.S. Air Force. He leads a team of personnel tasked with integrating resources and capabilities into the Air Force's overarching $140 billion annual plan. Quinn graduated from the University of Florida with a degree in aerospace engineering and holds master's degrees in business administration and national security studies. Quinn has commanded a pilot training wing, served as a fighter wing Vice Commander, commanded at the squadron level, and held staff positions at multiple levels. He has deployed in support of operations in southwest Asia and northern Africa.

Maj D. M. Smalenberger, USAF—is the Quantitative Augmentation Branch Chief on the Air Staff and is the Creator and Chief Architect of Project QuANTUM and Laniakea. He holds a BS in Operations Research from the USAF Academy, an MS in Operations Research from the Florida Institute of Technology, and a PhD in Operations Research from the Air Force Institute of Technology. He holds certifications as an Advanced Lean Six Sigma Master Black Belt and as a Project Management Professional. His research includes PAC-learnable systems, financial warfare, mathematical statistics, information theory, quantum-hardened encryption, and multi-valued logic.

Abstract

The impetus of this work is to develop a complete and ubiquitous approach to augment the planning, programming, budgeting, and execution (PPBE) decision process quantitatively. The proposed solution provides a methodology that emphasizes the ontology of senior leadership questions, highlights those data sets that are relevant, and leverages a suite of methodologies that meet the ontological requirements. We propose a data democratization and utilization platform (Laniakea) as well as a four-stage analytical engine (QuANTUM) to achieve this objective. We posit that tethering these approaches to the PPBE problem set will enhance the derived solutions through a technologically-scalable, mathematically-flexible, and factually-rigid solution and that this will enhance leadership's overall awareness, and to understand and chart a way toward cognitive supremacy.

Introduction

To achieve victory, an adversary with fewer resources must neutralize their disadvantage. This economic pressure will likely require an evolution in their strategy which maneuvers around these disadvantages. As the U.S. military retains its dominance over certain domains while pursuing others, it causes evolutionary pressures to its adversaries to adapt accordingly. Given the U.S. military's current trajectory, these evolutionary pressures are showing beyond traditional battle spaces, such as with Unrestricted Warfare (UW)[1], hybrid warfare, *sub rosa*,[2] or *casus fortuitus*[3] operations. To prepare for these types of

[1] From the Chinese phrase 超限战 and directly translates to "warfare beyond bounds." Its core tenets include financial and economic warfare, lawfare, information warfare, cyber warfare, and terrorism (Xiangsui & Qiao, 1999).

[2] Translated from Latin and means "under the rose." It is used to describe clandestine types of operations (Libicki, n.d.).

challenges, we must modify our approach to plan, program, budget, and execute with these in mind. By augmenting our decision processes quantitatively and with appropriate analytical techniques and technologies, we can attempt to achieve praxeological optimization toward cognitive supremacy to help counter these emerging and exotic types of threats as well as continue to hedge against classical sorts as well.

The intent of this initiative is to integrate scientific and analytical precision in the DoD's Planning, Programming, Budgeting, and Execution (PPBE) process while appreciating the ontological differences that the questions leadership propose may exhibit, demonstrating the power of authoritative data sources (ADS)[4] under correct conditions, and fully leveraging the full suite of analytical methodologies that exist. The PPBE process spans each domain and functional area, involves multiple stakeholders, and is constrained to limit objective information toward logical and feasible solutions. There is a critical need to provide pertinent facts, identify interdependencies, and understand historical rationale to ensure the highest probability of success while making investment and divestiture decisions for DoD force structure. A cause of these challenges listed above is the lack of adequate access to pertinent data to substantiate their decision process quantitatively due to the federated style of data management that currently exists, consequently causing inefficiencies in the "Data-to-Decisions" paradigm. The DoD requires a change to address this problem toward a technologically-scalable, mathematically-flexible, and factually-rigid methodology to replace the current monopoly via subjective interpretation.

We define the following four concepts specifically as they provide the necessary terminology for this paper: Cognitive Supremacy (CS), Quantitative Augmentation (QuA), Ubiquitous Modeling (UM), and Question Class (QC).

- *Cognitive Supremacy* is the ability to hold complete control in the decision space over an adversary.
- *Quantitative Augmentation* is a process which provides technologically-scalable, mathematically-flexible, and factually-rigid methodologies to enhance the decision process scientifically.
- *Ubiquitous Modeling* is the ability to describe any system using appropriate ontological methodologies.
- A *Question Class* is a set of questions which are ontologically[5] equivalent.

---

[3] Translated from Latin and means "an inevitable accident." It is used to describe non-reputable operations which may be carried out on an adversary. An example would be financial market manipulation and then attributing those effects to normal market fluctuations.

[4] "[A] recognized or official data production source with a designated mission statement or source/product to publish reliable and accurate data for subsequent use by customers. An authoritative data source may be the functional combination of multiple, separate data sources" (DoD, 2007).

[5] An ontology is the compartmentalization of problem solving techniques based on complexity, approach, computation, and variables that are required (O'Connor & Wong, 2012).

Question Classes

A typical methodology to problem-solving is to dissect it into its causes. Once we identify these causes, then we are in a position to influence the aggregate behavior by affecting one or many of these causes. Several process improvement initiatives over the years, such as Total Quality Management (TQM) and Air Force Smart Operations for the 21st Century (AFSO21), have provided helpful tools and mechanisms to do so. However, using their standard techniques implies that the problems we wish to understand in more detail and dissect into their causes lend themselves to a deconstructable approach. As we will illustrate, this assumption does not hold true for many PPBE-level types of problems, and so this strategy has conditioned the community to persist in an ill-fated strategy.

We propose an alternative to the typical process improvement problem solving by categorizing the types of questions based on their ontology. Once we establish a question's ontology, then we may apply those methodologies which meet the ontological assumptions. Figure 1 illustrates the alternative view of problem-solving based on the concept of an ontology.

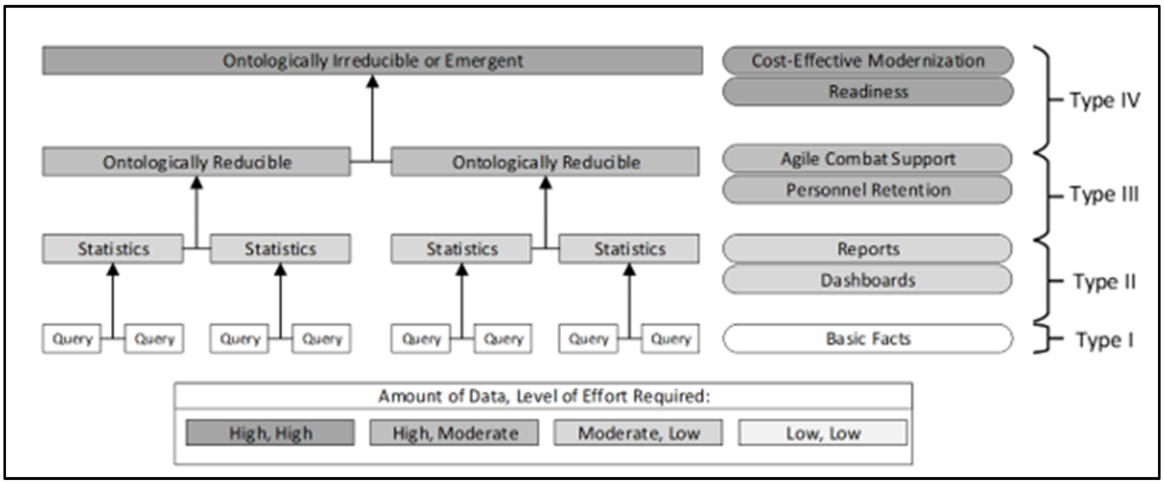


Figure 1. Illustration of the Hierarchical Structure for the Four Proposed QC With Pertinent USAF Examples to Provide the Reader Context: Personnel Recruitment and Retention (PR&R),[6] Agile Combat Support (ACS),[7] Readiness,[8] and Cost-Effective Modernization (CEM)[9]

---

[6] Represents the process that selects the personnel for the most appropriate Air Force Specialty Codes (AFSCs) based on their aptitude and preferences (Recruitment) and to optimize their dwelling within the enterprise (Retention).

[7] Represents the support tail of the USAF enterprise and comprises approximately 40% of Total Obligation Authority (TOA).

[8] The measure associated with efficiently allocating personnel and equipment to combat packets to meet operational objectives.

[9] A USAF priority to revamp aging aircraft, satellite, and nuclear fleets in a manner to maximize lethality and utility for a given cost structure.

Type I: Query

A query is a precise request for information retrieval from a database (DB) or information system (IS). The emphasis is on the content of the DB or IS and not on any ontology that may exist within the DB or IS. Queries require only the ADS on which we wish to run the query and require the least amount of technical expertise to do so. An example would be to list all females with the first name "Trista" from MilPDS.[10]

Type II: Statistics

A statistic condenses relationships which are stored in a DB or IS into parameters of a known statistical technique. There is an information transfer that occurs here from content to context. An example would be to test whether or not a normal distribution can accurately represent USAF entry test scores for enlisted members.

Type III: Ontologically Reducible Behavior

Deduced behavior based on the content of one or many DB or IS under consideration. We can dissect these relationships via one or several multivariate techniques and apply the typical cause analysis processes. An example would be how the fuel prices may constrain the number of sorties flown and may cause currency problems for fighter pilots.

Type IV: Ontologically Irreducible or Emergent Behavior

The resultant phenomenon where defined systems with clearly defined relationships may produce unpredictable and even unanticipated behaviors that are also more complex than any subsystem could produce in isolation. Therefore, we may not understand the etiological path as a set of few distinct and manageable factors. Furthermore, nearly all emergent-types of systems are ontologically irreducible to the lower levels (O'Connor & Wong, 2015). These systems include, but are not limited to, complex adaptive systems (CAS)[11] which many DoD systems and processes seem to be. This class will likely apply to all senior level questions that include "effectiveness," "agility," and "lethality."

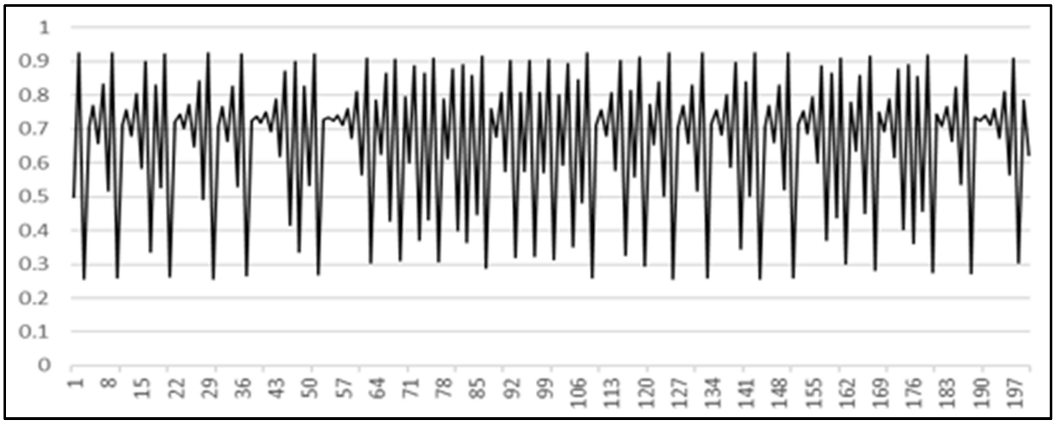
We propose a very simple example to illustrate how simple systems can become intractable to deconstruct. The output of a system we wish to model follows the logistic map in Equation 1, where x_n is a number between null and unity and λ represents a value between null and 4. This is a relatively simple system, and the future value is dependent on the prior value. However, we obtain very different behavior from this system depending on the value of λ that we choose. For some values of λ, this system will reach a steady-state; for other values, it will enter into oscillations, and for some values of λ, the system will enter a state with no clear pattern.

$$x_{n+1} = \lambda\, x_n(1 - x_n) \tag{1}$$

[10] "MilPDS is the single integrated 'Total Force' AF Human Resource system and authoritative data source for Total Force military records supporting all Active Duty, Guard, Reserve and retired AF members. MilPDS is the system of record that manages every aspect of an Airman's career, including accessions, assignments, career management, separation and retirement. MilPDS was the selected platform to realize the AF/A1 AF Integrated Personnel and Pay capability" (U.S. Air Force, 2017).

[11] A system in which a perfect information about each system subcomponent does not imply perfect information about the dynamics of all subcomponents in aggregate (Miller & Page, 2007).

As the analysts, we do not assume we know this relationship *a priori*, but wish to deduce this behavior from an ADS which captures observations from this system. Figure 2 illustrates an observation subset.



Note. We notice the erratic and unpredictable behavior that the system presents and which is difficult to predict or to deduce the underlying relationship analytically.

Figure 2. Illustration of the First 200 Observations From the Logistic Map With $x_0 = 0.5$ and $\lambda = 3.7$

Even with relatively simple systems such as in Equation 1, it may be difficult or impossible to deduce that these observations originate from a logistic map. A further complication is that the more complex the system is, the more observations you will likely require to validate a deduced relationship. However, there are likely insufficient observations necessary in a combination of ADS to sufficiently represent USAF questions of Type IV, and so we require a different approach which we will discuss later in this paper.

We posit that understanding and appreciating the class in which a question resides is a necessary but not sufficient step to articulating the question quantitatively. It provides a glimpse into the level of effort and the types of methodologies that you will likely require to answer it in a scientifically-rigorous way, and this is a focus later in this article. By understanding which class in which a question resides will help the analyst understand which types of approaches they will have to use to formulate the question in the Specific, Measurable, Attainable, Results-based, and Time-bound (SMART) framework (O'Neil & Conzemius, 2006). When taking the QC into account, then several Lean Six Sigma[12] techniques such as Ishikawa diagrams, 5 Whys, and so forth can assist an analyst in dissecting a question into pertinent and quantifiable sub-questions.

The query logic for Type I questions are a codified representation of the original question, and so as long as the data is available, this class is easy to answer. The answers we seek are only in the content of the data. For Type II questions, we seek the relationships, or the context, of specific data elements. Since the data content is available, we may run descriptive analytics against it and so arrive at specific statistics. These two QCs are nearly

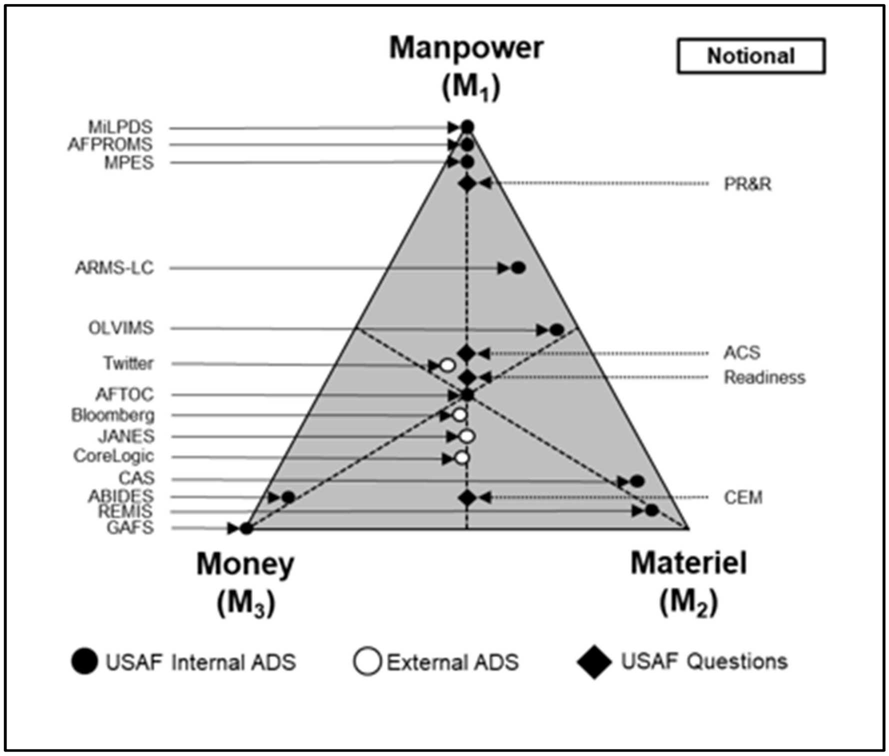
[12] A process improvement methodology which leverages collaborative team efforts to improve a system's or process's performance by removing waste and variation (Mills, Carnell, & Wheat, 2001).

trivial due to an analyst's ability to obtain an answer directly from the data. However, with Type III questions, an analyst is now interested in the interrelationships between disparate ADS. There may be some hidden behavior that an analyst may not understand or observe, and this may lead to missing ADS that we should have included or the introduction of an ADS into the solution that provides no additional utility. Type III questions will likely require an iterative process with question formulation and ADS selection. Lastly, there is no guarantee that any combination of ADS may sufficiently address Type III and IV questions. It is very easy to construct a scenario in which an infinite amount of data may not provide an adequate understanding of the underlying system behavior. For Type IV questions, we may be able to understand broad behavior, but it remains an open question if we may achieve a more granular level of understanding.

M^3-Space

The second component is to determine the fields in which ADS pertain to the question of interest. How to standardize and automate this approach remains an open question since there may be certain underlying behaviors present in the data that are not readily apparent and may involve fields that an analyst may overlook.

We propose a first step in developing a solution strategy. By analyzing many senior-level types of questions, a common theme becomes apparent. Many questions involve a combination of three core tenets: *Manpower, Materiel, and Money*, or M^3. Therefore, it seems advantageous to map each ADS (and then later map each field within each ADS) to the region between these core tenets to gain insight to which ADS may provide utility to certain types of questions once they, too, are mapped to the M^3-space. Figure 3 provides a notional example of several USAF core and external ADS and certain questions mentioned previously which are currently of interest to senior USAF leadership.



Note. The dashed lines represent the altitudes and represent locations where an ADS or question has equal representation for those core tenants on which it lies. Notionally, OLVIMS lies between M_1 and M_2 core tenants for example.

Figure 3. Illustration of the Proposed M^3-Space With Several USAF-Specific ADSs Such as GAFS,[13] MiLPDS, and REMIS;[14] Several External ADSs to the DoD Such as Bloomberg,[15] CoreLogic,[16] and Janes;[17] as Well as Several USAF Enterprise-Wide Questions

With this paradigm, we may chart the ADS locations between these three core tenets. This location will give us an idea of how likely an ADS may support a particular question of interest. Furthermore, if we map ADS which we have yet to obtain, then such an

---

[13] "General Accounting & Finance System (GAFS) is owned and functionally managed by DFAS-Columbus. It is used to process more than 3.2M accounting transactions totaling $3.4B monthly. GAFS-DTS processes more than$4.4M traveler payments annually; more than $4.5B in DoD travel payments" (U.S. Air Force, 2017).

[14] Reliability and Maintainability Information System (REMIS) is the primary USAF data system for collecting, validating, editing, processing, integrating, standardizing, and reporting equipment maintenance data, including reliability and maintainability data, on a global, world-wide basis (U.S. Air Force, 2017).

[15] One of the leading financial data vendors and brokers worldwide.

[16] CoreLogic is a data broker which specializes in U.S. economic, housing, and personnel data.

[17] Jane's Information Group was a British publishing company which specialized in military, aerospace, and transportation topics. It was acquired in 2007 by IHS Inc. which continues to sell their data products.

approach may assist in determining which ADS we should prioritize next. We will likely wish to have a balanced mix of ADS that may contribute to all three core tenets at any given time, and so this approach may assist in that objective.

Although the mapping of ADS is conceptually helpful, it only provides limited utility. Preferably, we require a granularity that is at the attribute level. The granularity will enable us to map each attribute to each core tenet or a combination of them. Unlike with the ADS mapping, an attribute mapping provides an added benefit that any attribute which maps to any altitude within the M^3 is a key ID across those core tenets. Furthermore, an attribute that maps to the centroid is a key ID for all three core tenets and would be of high value to tie multiple ADSs together.

Data Landscape

Data Policy

Standard practice for data analytics is to amass the necessary ADS locally or in the locally-owned and regulated environment. Each party signs some agreement to acquire one or more ADS to gain access or obtain a copy of one or more ADS. For DoD personnel and DoD organizations, this is usually via a Memorandum of Agreement (MoA) or via a System Authorization Access Request (SAAR) or DD2875. The MoA is a binding agreement between the parties involved and establishes which ADS the parties will access, how the customer will access, store, and use this data; what data protection measures they will enforce; and what their procedures will be to eliminate the data once the agreement has expired. Parties typically use an MoA when one or more customers request more than one ADS. Customers typically submit the variation DD2875 via their chain of command and then on to the specific ADS owner to gain approval to access or obtain a copy of a particular ADS.

Data Situation

A general observation is that ADS are heterogeneous over multiple operating systems (primarily Windows, Linux, and Solaris) and utilize several DB environments including but not limited to Actian, DB2, Ingres,[18] MS SQL Server, and Oracle as well as a variety of IS dating as early as the 1970s. In some cases, non-U.S. government third parties maintain these ADS outside of any DoD installation and are geographically disparate.

Observations and Challenges

The lack of a consistent DoD data acquisition process requires a considerable level of time investment to ascertain and comply with the nuances of any particular ADS. Typically, there is a low probability of establishing a relationship successfully with the owning organization if there is no potential for a symbiotic relationship or no higher authority mandates cooperation. A compounding factor to this challenge is that the DoD has little or no situational awareness about the number or location of DB or IS it possesses. Therefore, it can be difficult to request an ADS if it is arduous to locate the proper authority to request access. Furthermore, system administrators or database administrators (DBA) often are extremely specialized and therefore are no longer experts on an entire IS or DB. Therefore,

[18] The INteractive Graphics REtrieval System is an open-source SQL relational database which was developed at the University of California, Berkeley, in the 1970s and is still in use in the DoD today.

we may require several teams of individuals for a single ADS. A particularly concerning observation is the compliance with outdated data management policies. Some data owners retain data for a couple of years and then purge or overwrite them with new data. This practice is particularly troublesome since we are unable to provide meaningful analytics within a problem set that requires this data once this data no longer exists. The rationale for these policies was a cost savings measure when electronic storage costs were considerably higher than only a few years ago. However, it is not sufficient to gain access or to obtain a copy of an ADS to conduct a meaningful analysis. One potentially overlooked, yet critical, item is a legend or explanation of the attributes and values contained in any ADS—or data dictionary. There are many instances where data dictionaries are either outdated, incomplete, or in some cases nonexistent. Lastly, there is a systemic lack of continuity to explain the changes a DB, IS, or ADS have undergone since their inception. The lack of continuity provides difficulties in interpreting changes within these systems or data sources and which may have significant implications on the analytical quality which an analyst can produce.

Proposed Solution

We offer a potential strategy to address these challenges by bifurcating, yet tethering, the problem into a data approach and analytics approach, as shown in Figure 4. To provide high-quality analytics that are useful to the larger community, we require a standardized process and framework to access and leverage many ADS. Since ADS within the DoD are heterogeneous, we require a considerable level of effort to Extract, Transform, and Load (ETL) them into a common environment. To leverage a large number of feasible tools, a cloud environment in the private sector is likely an improved solution over purchasing a suite of licenses for different ETL tools. A cloud solution would enable the ETL process to utilize tools that may also not possess the requisite approval for installation on machines connected to a U.S. government network. The Enterprise Information Model (EIM) provides the aggregation of knowledge for all data dictionaries, their commonalities, and combinations of cross-ADS attributes which analysts have used in the past. The environment that enables the ETL process, storing of ADS, and provides an overarching cloud solution for the AF and possibly DoD, is Laniakea.[19]

We also suggest a common analytical base for the larger DoD community. The original intent of this analytical base was to have the ability to access and utilize ADS within Laniakea easily and so provide a more useful analytical toolset to the communities which wish to conduct analytics on a wide span of ADS. As previously argued, any analytical approach must provide sufficient flexibility to address each of the four types of QC. The platform that provides accessibility to Laniakea while providing a suite of analytical tools to the DoD community is the Quantitative Augmentation via Neuro-evolutionary Technologies toward Ubiquitous Modeling or QuANTUM.

---

[19] The word *Laniakea* is Hawaiian and means ``immeasurable heaven.'' It is the galaxy supercluster which contains the Milky Way and an estimated 100,000 other galaxies and spans approximately 520 million light years. In this context, Laniakea refers to the USAF's galactic data supercluster which seems appropriate due to the vastness of the USAF's and DoD's ADS.

The solution connects to the DoD Information Network (DoDIN)[20] via an Internet Access Point (IAP), internet gateways, and Virtual Private Clouds (VPC). Each component, the web application firewalls, the bastion hosts, Laniakea, and QuANTUM, are physically and virtually distinct entities which connect with VPC peering which would enable added levels of security and inhibit transitivity of access to each isolated component in the chance an adversary exploited a vulnerability. Also, both Laniakea and QuANTUM have persistent monitoring to protect against threats or anomalous activity actively.

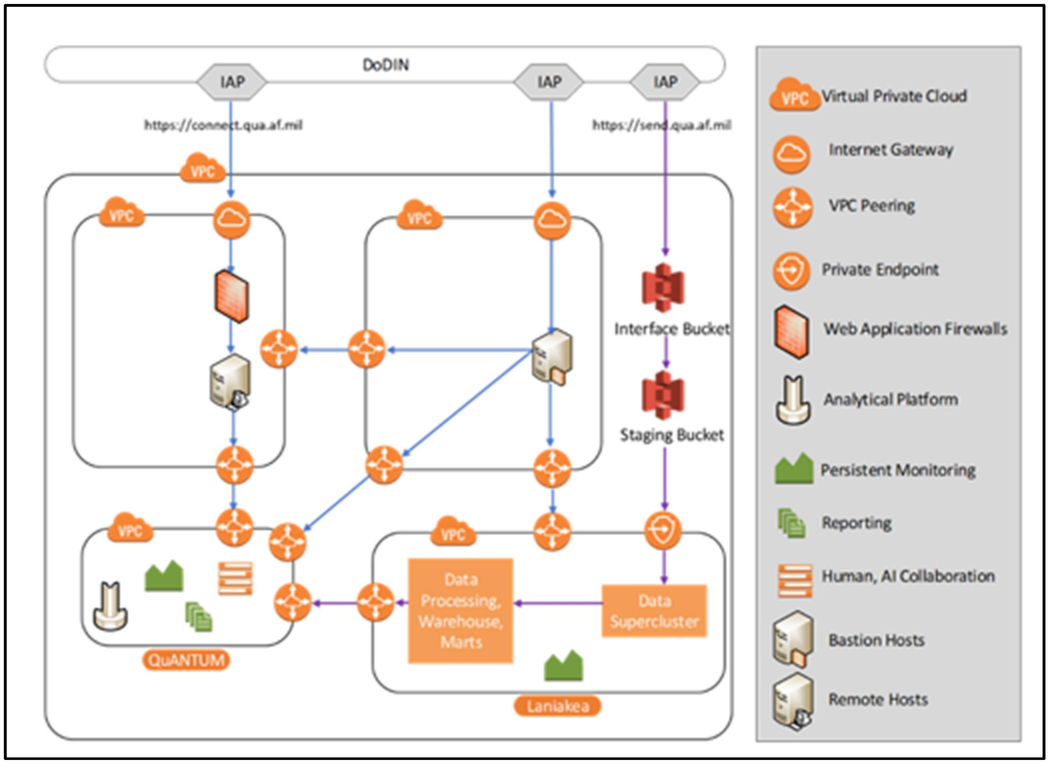


Figure 4. Illustration of the QuA Ecosystem as Conceived by the Authors

[20] "The set of information capabilities, and associated processes for collecting, processing, storing, disseminating, and managing information on-demand to warfighters, policy makers, and support personnel, whether interconnected or stand-alone, including owned and leased communications and computing systems and services, software (including applications), data, security services, other associated services, and national security systems" (Committee on National Security Systems, 2003; Joint Chiefs of Staff, 2010).

The Analytical Engine—QuANTUM

We require an analytical approach which would enable us to leverage the scientific process and, concerning the questions above, classes and ADS since there are multiple types of QCs and these classes require different strategies to address them adequately. Providing a common platform to access ADS and analytics is preferable since it would provide the opportunity to develop a common terminology throughout many communities. We posit that one of the difficulties in explaining a viewpoint is that the terminology is defined in disparate colloquialisms and human language and not on mathematics.

We illustrate the example of Fully-Burdened Cost (FBC). The first challenge to using a definition is to locate one that is accepted by the larger DoD community. The definition for FBC seems to vary based on Service, organization, and experience. Typically, the Government Accountability Office (GAO) publishes lists periodically of general terms for the PPBE and other processes (GAO, 2005). There are three types of FBC listed in the Financial Management Regulation (FMR), namely cost for fuel, the unit cost of contractor-acquired property, and composite rate of pay, allowances, taxes, and accruals (Under Secretary of Defense, Comptroller, n.d.). Each of these definitions is not quantitatively well-defined, and so the DoD guides estimate these types of fully-burdened costs (DoD, 2013). Since the DoD only provides not completely quantifiable guidance on how to interpret these definitions, it leaves the opportunity for variability in the calculation of these costs. This variability may provide confusion and inconsistencies in planning, programming, budgeting, and execution, and these inconsistencies may compound and result in suboptimal choices. An alternative approach would be to define these terms using a query language, such as Structured Query Language (SQL) that can access many records within ADS and so calculate a value that we can tether to fact. This proposed approach would incentivize many DoD communities which use similar terms such as FBC to develop common definitions so they may communicate with one another more intelligibly. This unprecedented clarity would be of additional benefit to higher levels of leadership, such as the Office of the Secretary of Defense (OSD) since it would no longer need to interpret many inconsistent definitions of the same terminology.

The next logical step after creating a standard data repository, such as Laniakea, and providing quantitative definitions for common terminology, would be to provide an analytical environment that may augment the decision process quantitatively and for varying QCs. The typical approach to conducting analysis is to amass those necessary ADS which are most pertinent to answering a question locally and then to devise an approach with those analytical tools which an analyst has at their disposal. Often the analyst is limited to the types of analytics they can perform since they have very few tools, and so they must create a toolset. This limitation is typical since many organizations cannot afford analytical suites of tools that would allow them to answer their questions more swiftly and rigorously. Also, the authorities for DoD IS have not yet cleared many formidable tools for installation on computers connected to government networks. This primary approach to conduct analytics is no longer feasible given financial, security, and other constraints. Therefore, we posit that an alternative solution exists that provides a higher level of accommodation for these concerns. We observe that the primary approach involves aggregating ADS to where the analytics reside. The alternative approach, or dual, would push the analytics to where the ADS reside. This alternative would enable an analyst to generate a virtual machine (VM)[21] in the cloud solution we mentioned and to virtually install an instance of many types of analytical products that would not be allowed otherwise. Although this provides an improved solution, the difficulty remains with having a sufficient number of licenses or proprietary products for the communities.

An improvement would be an analytical platform which functions as a Platform-as-a-Service (PaaS) and utilizes the ADS stored in Laniakea as a Data-as-a-Service (DaaS) approach. Therefore, this platform would function as a utility such as water, electricity, natural gas, or internet. This framework leverages economies of scale in that the cost per user declines as the number of users rises. Therefore, the incentive is for the developmental teams to maintain a viable and demonstrably superior platform for the communities to use. Therefore, this paper proposes a Government-Off-The-Shelf (GOTS)-owned platform that leverages a suite of ADS, provides common terminology via a query language, and provides a suite of analytical methodologies from which to choose. The platform takes advantage of computing at multiple classification levels and the ability to scale with its compute infrastructure.

QuANTUM is an analytical engine that is comprised of four stages and is in parallel development with Laniakea as its data source. Each stage addresses and enhances the efforts of the prior stages and addresses their potential limitations. Stage I, Argos,[22] is the analytical platform that provides a suite of tools ranging from simple queries from ADS to multivariate and advanced approaches such as topological data analytics (TDA), machine learning (ML), forecasting, and optimization. However, Argos can only address Types I through III. The reason is that we require historical observations to address these questions. With historical data, Argos is incapable of adequately addressing Type IV questions due to those reasons we discussed.

To address the Type IV QC, we require an ontologically distinct methodology from that of Type I through III, and this difference is the focus of Stage II, Krishna.[23] Since this is an approach to address the challenges as discussed briefly with the logistic map, we require a methodology that would provide some level of insight that lies beyond both internal and external ADS but can still leverage these ADS when appropriate. One possible methodology would be to utilize modeling and simulation (M&S) and so attempt to gain insights into the emergent behavior that may exist as simpler systems interact with one another. We may wish to compute many simulations with different initial conditions to get a general sense of how the system under consideration may behave. If there are significant deviations from the baseline system, then this might provide evidence to suggest that the system may require additional modifications to align both real and simulated systems more closely. The M&S environment that the Type IV QC concerning the DoD would require is a physics-based, continuous-time, imperfect information multi-agent framework at the campaign or multi-campaign level. A physics-based approach is necessary to provide realism to the

[21] An emulator of an IS. It is a software substitute for a physical machine.
[22] Also known as Argus, Panoptes is a many-eyed, all-seeing giant in Greek mythology
[23] The eighth avatar of Vishnu. He is the central figure in the Mahabharata, Bhagavata Puana, and the Bhagavad Gita. The impetus of the name is due to verse 32 in the Bhagavad Gita which also drew Robert Oppenheimer’s attention: “Now I am become death, the destroyer of worlds” (Mascaro, 1962).

environment. The model includes the application of physical laws and constraints within the environment such as gravity for munitions or aircraft and Line of Sight (LoS) for communication links. The feature of continuous time is important since complex systems may behave very differently when discretized. Also, providing the correct approach to the flow of time will provide an improved model concerning other types of flows (e.g., information) within the framework. A common feature many wargames fail to appreciate fully is the limitation of timely and perfect information about the situation on the battlefield, and is analogous to von Clausewitz's (1832) "fog of war." By limiting the fidelity and flow of information, we introduce risk into the decision space and thereby provide the framework to learn how to operate in these degraded conditions. The intent is to introduce these features in an environment that remains scale-invariant and therefore provides utility to leadership at any level. Lastly, and perhaps most importantly, is the introduction of allied and adversarial adaptive systems. We may begin to develop this environment by utilizing Reinforcement Learning (RL)[24] and then exploring the utility of Generative Adversarial Networks (GAN).[25] These types of approaches provide evolving solution strategies given a complex environment and with minimal rules provided *a priori*. Also, these types of approaches may find novel ways to interact with a complex environment that someone may not anticipate. Those types of strategies which exploit indirect and cascading effects are those which the authors feel may be of importance in future conflicts and require further research to explore adequately.[26] If we provide this system flexibility with determining which assets to locate where given a certain set of campaigns, we may also gain insights into how our asset portfolio should be allocated to achieve the highest probability of victory given those campaigns (or any other objective function we choose). Finally, we wish that this framework exploit advances in massively parallel computational platforms and that these methodologies may learn from many environments simultaneously, thereby providing a richer awareness of the solution space to leadership in less time.

Krishna is no panacea. One of its major limitations is that it provides feasible strategies given the current weapon systems and infrastructure that exist. Allied and adversarial weapons portfolios are dynamic, and systems with enhanced, even new, capabilities arrive with some regularity. Stage III, or Oracle,[27] provides the framework necessary to address Krishna's limitations. It introduces degrees of freedom (DoF) to those weapon system specifications (e.g., maximum range, maximum altitude, maximum munitions capacity, etc. for aircraft, and so forth) that may change with new weapon systems or modifications of older systems. However, any analysis which provides modifications to

---

[24] One of the three main types of learning: supervised, unsupervised, and reinforcement learning. Over successive iterations, an RL model will learn based on the environment in which it is introduced. If the aggregate behavior of the RL system is desirable, then it is rewarded and so reinforced otherwise it is not (Russell & Norvig, 2015).

[25] Two unsupervised artificial neural networks which contest in a zero-sum test space (Goodfellow et al., 2014).

[26] A possible scenario may involve targeting several seemingly ancillary and unguarded power substations simultaneously. Their cumulative effect may introduce a rolling blackout which may degrade an adversarial military installation communications network and so degrade the adversary's command and control.

[27] Named after the Oracle of Delphi and not to be confused with the company that bears the same name.

specifications alone is inherently flawed, as we would also need to consider the trade-offs which we incur with these modifications and how these new or modified systems should interact with the rest of the portfolio in a given a set of campaigns. One of the main trade-offs is that we incur Research, Development, Test & Evaluation (RDT&E) costs for any given modification. These need to be captured into the specification changes as well.

We may observe that in this framework there may arise common strategies and themes within the behavior of the systems that may confirm or repudiate a commonly-held belief. These observations would provide a level of substantiation not yet reached via alternative methods. The intent would be to augment policies and strategies at the Service, Joint, Department, and coalition levels where and when appropriate. Furthermore, the advantage of this approach is that it does not rely on the analysis of the levels above to obtain a solution. This approach might be advantageous since each of these levels attempts to achieve optimality within their sphere of influence, and this effect might result in many local optima rather than an optimum solution for all actors at the same time.[28] Referring to Type IV types of questions, models of subsystems may also not provide an accurate description of aggregate behavior, and therefore an overarching model may also be more appropriate here.

Unlike Argos and Krishna, Oracle is a composite stage, and therefore Oracle draws on specific capabilities that the prior stages deliver. We may leverage certain methodologies incorporated into Argos to develop a set of relationships between the cost of a certain specification set and attempt to infer an overall relationship between cost and specifications. We define this as the COst-to-Specifications Manifold (COSM). Although we may apply multivariate techniques such as Response Surface Methodologies (RSM)[29] to capture this relationship accurately, there is a growing trend to leverage alternative techniques such as Artificial Neural Networks (ANN)[30] over RSM (Himmel & May, 1991; Carpenter & Barthelemy, 1993; Hussain, Xuanqiang, & Johnson, 1991). This realization enables us to recycle techniques present in Argos and utilize them for Oracle. Figure 5 provides an overview of the process discussed in this section.

---

[28] We can never guarantee reaching a global optimum when utilizing heuristics, and so the argument is that we need to traverse the solution space given all actor constraints simultaneously.
[29] The approach to utilize mechanistic models, empirical, and response surface models to identify and fit factors and experimental data to an appropriate mathematical representation (Myers, Montgomerey, & Anderson-Cook, 2009).
[30] A synthetic analog to biological learning which is also robust against errors and incomplete input (Mitchell, 1997).

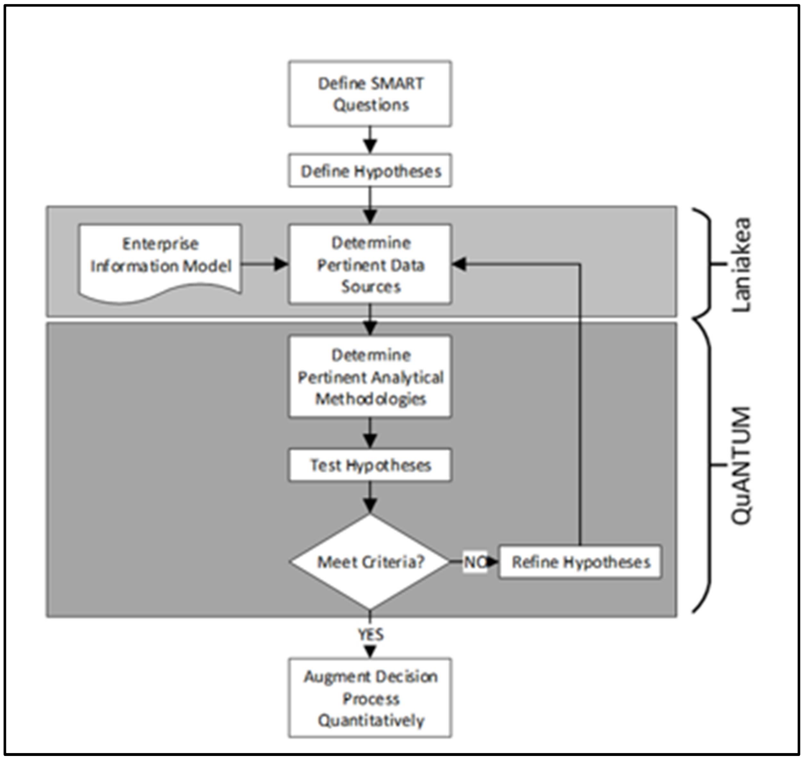


Figure 5. Illustration of the Scientific Process Utilizing the QC to Help Frame SMART Questions and to Shape Hypotheses While Leveraging the Galactic Data Supercluster, Laniakea, and the Analytical Engine, QuANTUM

Once we have established an acceptable COSM model, we can then provide some insight into the trade space between these two factors. The COSM functional approximation now provides the necessary cost penalty to any proposed improvement by an M&S solution, such as Krishna. As the agents within Krishna hypothesize what improvements to make to existing systems, they will incur the necessary cost penalties of making certain specification modifications. Therefore, their resultant strategies will account for these nuances as they try to achieve their overall objectives.

No change can happen instantaneously to a weapons plan, strategy, or weapons portfolio. These changes will require appropriate planning for these modifications, reprogramming funds, budgeting resources within future years, and executing these resources in a manner that remains feasible within political, policy, fiscal, and other types of constraints. We therefore require feasible scheduling trajectories for these improvements given their constraint set. This problem is a classic resource-constrained weighted scheduling optimization problem (RCWSOP) for which Genetic Algorithms (GA),[31] among

[31] A synthetic analog to Darwinian evolution at the genetic level. Information is encoded within artificial chromosomes and scored based on a supplied objective. As in biology, these chromosomes undergo crossover and mutations to provide variety and so traverse the solution space more completely. The chromosomes which provide the best fit are likely to carry that information on into successive generations (Mitchell, 1998).

other heuristics, have been used successfully to provide acceptable solutions. These approaches are highly parallelizable, and so if we supply additional computational resources, we may traverse a larger portion of the solution space and so attempt to find and improve on those feasible solutions we have obtained. Since the RCSOP is Type III ontologically, we may define Kronos as a derivative stage based on methodologies within Argos. Figure 6 provides a detailed overview of the QuANTUM analytical engine.

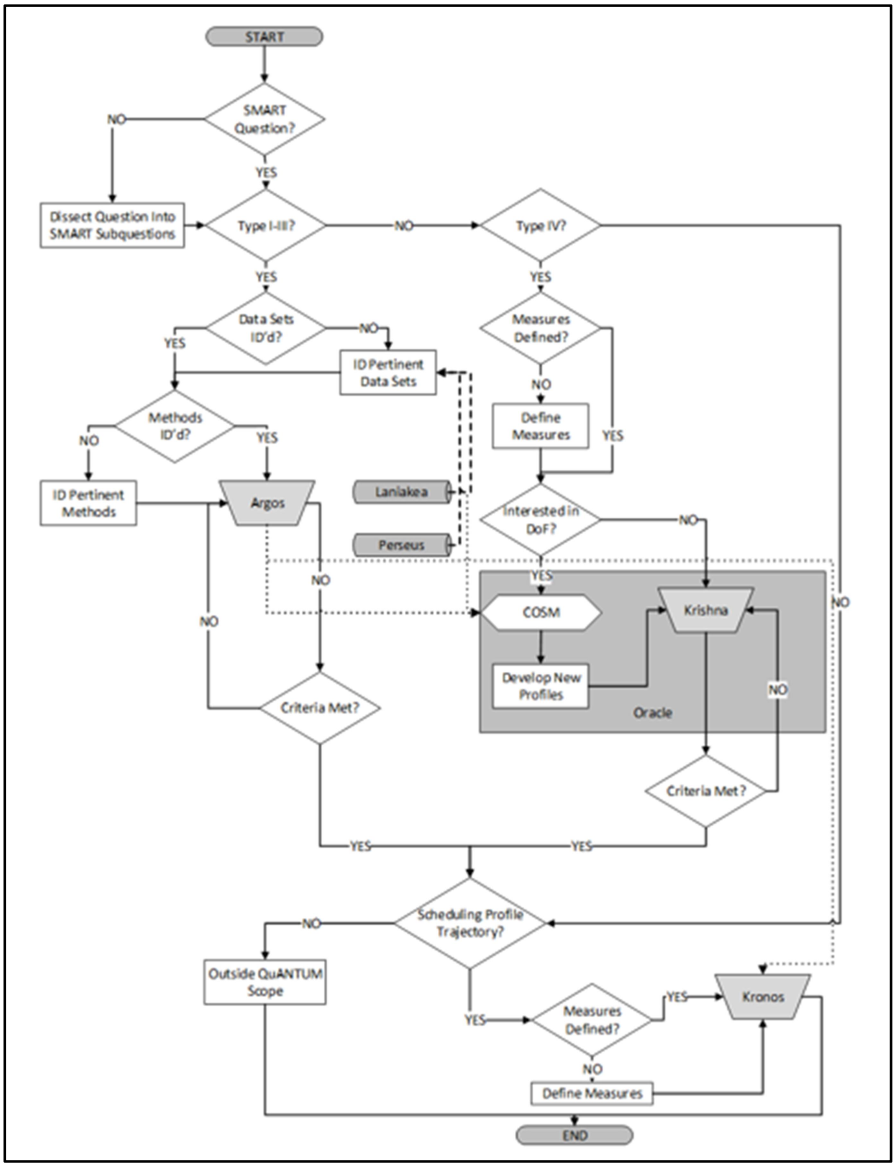


Figure 6. Proposed QuANTUM Analytical Engine Process

The Galactic Synthetic Data Supercluster—Perseus[32]

Currently, most DoD Data Democratization (DoD^3) and analysis efforts face limitations due to the nature of data (or observation) classification. The impetus of these classification guidelines is to protect data specific to an individual, such as Personally Identifiable Information (PII) and Personal Health Information (PHI), as well as data about national security which exists at levels such as secret and top secret. Proper automated and consistent classification methods of aggregated ADS remain an open question and an active field of research. The lower bound of a classification process is seemingly trivial, as any aggregation would inherit the highest classification level of the individually aggregated ADS. However, it is unclear how to determine the upper bound at aggregation, as well as after the implementation of specific analytical methods on those ADS. Therefore, the analytical communities require a feasible way forward as an interim solution as the defense community continues to tackle this question.

The classification of those ADS post-aggregation depends on their information. DB and IS store information in the content and the encoded in the relationships within and between those ADS. Ergo, if we can regulate the information of the ADS, then we may adjust their classification levels as necessary. Different classification levels require different levels of emphasis within these two core tenets. Specifically, about PII and PHI-types of data, any method must eliminate the possibility of direct targeting of an individual if the data set remains intact. However, it is not necessary to eliminate relationships within the data as these amongst individuals are not PII or PHI. About operational data, there are circumstances where the content of the ADS, as well as the relationships within the ADS may be at the same or higher classification level, and so both tenets require consideration. We assume the following:

1. The utilized ADS are sufficiently complete and of sufficient dimension to imitate given some methodology.
2. The underlying ontology is Probably Approximately Correct (PAC)[33]–learnable.

About Assumption 1, to effectively use real observations to derive synthetic observations, we require observations that are sufficiently complete to do so. We assume that the data is usable and does not require considerable preparation, interpolation, or any other approach for its use. Many of the ADS within the DoD have not yet reached a sufficient level of completeness to derive synthetic observations. Furthermore, and based on prior data extraction observations, many DoD data sets do not necessarily possess sufficient depth for the level of width. Therefore, there will likely be challenges with the generation of synthetic data if there are a sufficient number of observations available for their ontology. We posit that most systems are likely to exhibit highly nonlinear behavior and so would require many more observations to imitate the underlying behavior better. About

[32] The Perseus-Pisces supercluster (SCI 40) stretches approximately 250 million lightyears and is one of the largest known structures in the nearby universe, while the other is Laniakea, in which the Milky Way resides.

[33] Proposed by Leslie Valiant in 1984 and is an aspect in machine learning which provides a framework for mathematical analysis. The objective is to define a system that has low generalization error of the selected generalization function (or hypothesis) which is bounded in polynomial time.

Assumption 2, we assume that there is sufficient ontological information encoded within the real observations under consideration. This assumption would imply that some methodology exists which can approximate the content and context within the data sets under consideration.

This capability is exceedingly important, as aggregation will likely increase the overall classification level and so will likely diminish the number of organizations which may be privy to using this data. In many cases, the DoD relies on academic institutions, government contractors, or other government organizations or agencies to enhance their understanding and so increase their potential capability. Providing a mechanism to pre-select how closely synthetic observations may imitate real observations would enable an organization to retain these relationships without having to sacrifice aggregation. The underlying mechanism to generate synthetic observations in a meaningful way and the necessary halting criteria can be found in other works (Smalenberger, 2018).

Data Security—A Flawed Concept

Systems that contain or transmit data (and so transmit information) have always been prone to attack. We may attribute attacks to the fourth general principle of economics: "People respond to incentives" (Miller, Benjamin, & North, 2002). The typical approach has always been to focus on enhanced security features, including physical security, firewalls, limiting access to the data, and more exotic encryption techniques. Each of these approaches ultimately fails, and it is only a function of time until they do. It seems likely that a better version of the same old approaches is insufficient to diminish the incentive structure for the adversary.

These older strategies originated when data storage and processing on that data was at a premium. Therefore, the only data you wished to store was that data which was meaningful. Any user would delete the rest. Today, however, data storage and IS are orders of magnitude cheaper than they were previously. We propose a strategy to take advantage of that.

Assume that the synthetic observations stored in Perseus are not degraded in any form to provide them to academia, etc., but imitate the observations stored in Laniakea in a way that is no longer distinguishable between the two. If we derive the entire Perseus data supercluster in this way, we could generate a fully synthetic version of Laniakea within Perseus. This tactic would mean that given both data superclusters, it would not be possible to distinguish between the two which one contains real observations and which one contains synthetic observations. However, this approach still provides a large incentive for an adversary since a 50% chance remains that they select the correct data supercluster. Also, they could select both and provide contingency plans in the 50% chance that they selected the incorrect one.

A possibly improved approach would be to introduce the synthetic observations into Laniakea at tracked but seemingly random locations. Using this approach, we may introduce a much larger number of synthetic observations into Laniakea that remain indistinguishable from the real observations. Therefore, it would be increasingly difficult to know which observations are real and which observations are synthetic to exploit them. This stenographic[34] approach limits the incentive structure of an adversary.

However, the tradeoff with using this alternative approach is that certain attributes will need to be flagged and duplicated across the synthetic data fields as well. These are typically those key fields such as a person's name, social security number, and so forth. Once these are duplicated across the synthetic observations as well, it will be extremely arduous to classify real from synthetic observations. Additional research needs to be conducted to validate this as a feasible approach.

Summary

Due to adversarial evolutionary pressures, the DoD requires an improved approach to conduct the PPBE process which is technologically-scalable, mathematically-flexible, and factually-rigid. The proposed methodology in this paper provided a potential approach to determine the ontology of important senior leadership questions, develop a strategy to map these to pertinent ADS via M^3 and Laniakea, and leverage analytical techniques that meet their assumptions and solution requirements via QuANTUM. The paper also indicated how someone might generate synthetic data via Perseus which would enable communities to participate in the platform's future development while addressing classification concerns with ADS aggregation. Lastly, the paper addressed a potential approach to addressing the critical flaws with data security by using stenographic encryption.

References

Carpenter, W. C., & Barthelemy, J.-F. M. (1993). A comparison of polynomial approximations and artificial neural nets as response surfaces. *Structural Optimization, 5,* 166–174.

Committee on National Security Systems. (2003). *National information assurance (IA) glossary* (CNSS Instruction 4009). Ft. Meade, MD: Author.

DoD. (2007). *Unique identification (UID) standards for a net-centric Department of Defense* (DoD Directive 8320.03). Washington, DC: Author.

DoD. (2013). *Estimating and comparing the full costs of civilian and active duty military manpower and contract support* (DoDI 7041.04). Washington, DC: Author.

GAO. (2005). *A glossary of terms used in the federal budget process*. Washington, DC: Author.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, J. (2014, June 10). Generative adversarial networks. Retrieved March 11, 2018, from https://arxiv.org/abs/1406.2661

Himmel, C. D., & May, G. S. (1991). Advantages of plasma etch modeling using neural networks over statistical techniques. *IEEE Transactions on Semiconductor Manufacturing, 6,* 103–111.

Hussain, A. S., Xuanqiang, Y., & Johnson, R. D. (1991). Application of neural computing in pharmaceutical product development. *Pharmaceutical Research, 8,* 1248–1252.

[34] Stenography is the process of concealing a data set within another data set.

Joint Chiefs of Staff. (2010). *Joint Communications System* (JP 6-0). Washington, DC: Author.

Libicki, M. C. (n.d.). *Sub rosa cyber war*. Santa Monica, CA: RAND.

Mascaro, J. (1962). *The bhagavad gita.* Penguin.

Miller, J. H., & Page, S. E. (2007). *Complex adaptive systems: An introduction to computational models of social life*. Princeton, NJ: Princeton University Press.

Miller, R. L., Benjamin, D. K., & North, D. C. (2002). *The economics of public issues*. Boston, MA: Pearson Education.

Mills, C., Carnell, M., & Wheat, B. (2001). *Leaning into Six Sigma: The path to integration of Lean Enterprise and Six Sigma.* Boulder City, CO: Publishing Partners.

Mitchell, M. (1998). *An introduction to genetic algorithms*. Cambridge, MA: MIT Press.

Mitchell, T. M. (1997). *Machine learning*. New York City, NY: MIT Press & McGraw-Hill.

Myers, R. H., Montgomerey, D. C., & Anderson-Cook, C. M. (2009). *Response surface methodology: Process and product optimization using designed experiments*. Hoboken, NJ: John Wiley & Sons.

O'Connor, T., & Wong, H. Y. (2012, February 28). Emergent properties. Retrieved March 12, 2018, from https://plato.stanford.edu/archives/spr2012/entries/properties-emergent/

O'Connor, T., & Wong, H. Y. (2015, March 21). Emergent properties. Retrieved March 1, 2018, from https://plato.stanford.edu/archives/sum2015/entries/properties-emergent/

O'Neil, J., & Conzemius, A. (2006). *The power of SMART goals: Using goals to improve student learning.* Bloomington, IN: Solution Tree Press.

Russell, S., & Norvig, P. (2015). *Artificial intelligence: A modern approach*. Harlow, England: Pearson Education.

Smalenberger, D. M. (2018, forthcoming). On the imitation of real observations via statistical collisions for informational fracking (SCIF).

Under Secretary of Defense, Comptroller. (n.d.). *Department of Defense Financial Management Regulation* (DoD 7000.14-R). Washington, DC: Author.

U. S. Air Force. (2017). *Life cycle management center business enterprise systems reference guide.*

Valiant, L. (1984). A theory of the learnable. *Communications of the ACM, 27*(11), 1134–1142.

von Clausewitz, C. (1832). *Vom kriege.*

Xiangsui, W., & Qiao, L. (1999). *Unrestricted warfare*. Beijing, China: PLA Literature and Arts Publishing House.

Disclaimer

The views expressed in this document are those of the authors and do not reflect the official policy or position of the United States Air Force, the United States Department of Defense, or the United States Government. This work is academic only and should not be used to imply or infer actual mission capabilities or limitations.

ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL
555 DYER ROAD, INGERSOLL HALL
MONTEREY, CA 93943

www.acquisitionresearch.net