Analytic Technology Industry Roundtable

Analytic Technology Industry Roundtable Study: Architecture Survey and Review

Dr. Ransom Winder, MITRE

Mr. Joseph Jubinski, MITRE

Dr. Angela O'Hanlon, MITRE

November 2016

Analytic Technology Industry Roundtable

This page intentionally left blank.

Introduction

Analytics have become commonplace and essential throughout government and industry. While integration across analytic capabilities and results is poised to provide deeper insights as data's scale and complexity grow, isolated proprietary models of workflow and results hinder collaboration. This study offers an overview of the landscape of available and favored analytic architecture solutions across government agencies, both established practice and emerging capability, giving the necessary context for a reference implementation for an analytic architecture enabling collaboration between government and industry. This study also examines solutions for specific elements of an exchange service supporting such an analytic architecture, with an emphasis on those that are readily available.

Analytic Architecture Overview

Analytics are software applications and services that discover meaningful and worthwhile patterns in data. While many analytics can function in a stand-alone manner, they can potentially be applied in larger workflows for deeper understanding of data that is first transformed, reduced, or enriched into interim results by upstream analytic processes. This suggests there is value in defining an analytic architecture that supports these potentially complex workflows. From this perspective, it would be ideal if the data, the analytics, and the analytic results were all equally accessible when building the architecture at the outset, but these elements are components that are often proprietary or are core intellectual property that allows an organization to hold a competitive edge, even if the organization might be willing to provide or sell results. A successful architecture intending to foster collaboration between government and industry partners should account for this in how it handles workflow, scalable storage, and modeling the data and resulting knowledge.

Analytic Architectures in Government

Our examination of government practices with analytic architectures explored a wide variety of government organizations. The following subsections describe some of the significant existing and emerging analytic architectures identified across the different government agencies, as well as additional context as it helps to clarify each situation.

Defense and Intelligence Enterprises

The defense and intelligence communities in recent years have been trying to shift technology to allow for a greater capacity for data sharing among analysts, the efforts of which have been captured in initiatives such as the Intelligence Community Information Technology Enterprise (IC ITE), which emphasizes integration of classified data and permitting workflow between intelligence agencies, and the Department of Defense (DoD) Joint Information Environment (JIE), a broader effort that looks beyond intelligence to span secret and unclassified levels and handle logistics, operations, and other tasks [1].

Particularly noteworthy in this context is the Defense Intelligence Information Enterprise (DI2E) framework, which seeks to integrate all DoD and intelligence information and tools, also effectively bridging IC ITE and JIE [1, 2]. With these enterprises and an underlying cloud technology, the capacity to collect, store, and distribute data at scale is possible, but they also open up the potential for automated analytics and analytic frameworks to be applied to enrich the data, mitigating or supplementing historically manual processes in some cases.

A key goal of IC ITE is to provide broader and standardized access to analytic applications and tools [3]. As such, within IC ITE, there is an IC Applications Mall, which consists of a development environment that facilitates the trading of application code and a hosting environment where applications can be uploaded and downloaded, all to ensure that important applications can be re-used without having to be wholly re-engineered from scratch [4, 5]. In DI2E, a similar exchange for application sharing exists [1], but perhaps most significantly DI2E is composed of software services that allow elements such as the Distributed Common Ground Systems (DCGS), which we discuss further below, and Combatant Command Joint Intelligence Operations Centers (COCOM JIOC) to operate as a global enterprise [6, 7].

Distributed Common Ground System

A significant component of the enterprises discussed above, the Distributed Common Ground System is commonly used to refer to many evolving systems, such as those for Air Force (AF DCGS), Army (DCGS-A), and Navy (DCGS-N).

For the Air Force, this is their primary intelligence, surveillance, and reconnaissance (ISR) weapons system for planning and direction, collection, processing and exploitation, analysis and dissemination [8]. This effectively means it is the system by which intelligence products are created and distributed both within their forces and to their mission partners. AF DCGS has the capability to have a global reach and near-real-time speed of intelligence to different theaters of operation across its communications architecture, bridging ISR collection platforms with intelligence analysts and the leadership to allow for data to drive decisions [9].

Recent advances in AF DCGS seek to move it toward a more open and agile architecture, with a plug-and-play environment for applications allowing for faster deployment [10], and the long term plan [11] includes providing automated support and applying machine learning techniques to alleviate certain manual tasks and to allow analytic tools to draw from different sources.

Similar DCGS solutions also emerged for Army and Navy (DCGS-A and DCGS-N, respectively), where the analytic architecture pattern is apt. Furthermore, the DCGS Integration Backbone (DIB) serves to allow DCGS solutions from the different service branches of the U.S. armed forces to network with one another, facilitating the sharing of intelligence [12], as depicted in Figure 1. Also relevant in DCGS, as exemplified in AF DCGS, is the need to handle large volumes and high velocities of data, with many terabytes per day, often of large image and video files in the case of AF DCGS [8].

Much as AF DCGS is the Air Force's key ISR system, DCGS-A is the Army's parallel effort for producing and disseminating intelligence, both within and beyond their service branch [13, 14]. As DCGS-A continues to evolve, analytics and user-interfaces for analytic results are a priority [15] and may be integrated, relying on improvements in cloud computing and analytics for big data to achieve, among other things, scalability, usability, reliability and availability. At present, the Army is considering the Tactical Cloud Reference Implementation (TCRI), a tactical cloud platform for ISR capability [16, 17], described further below.

As with other DCGS instances, the Navy's DCGS-N emphasizes data fusion and intelligence dissemination, and there is a significant analytic architecture under development and being evaluated as part of its Increment 2 effort. As noted with DCGS-A, this framework is the Tactical Cloud Reference Implementation [18, 19]. In order to fulfill the DCGS-N requirements, this implementation must span the Navy, both afloat and ashore, as well as having the capacity to share with other intelligence services as well. As its name suggests, TCRI is built on underlying cloud technologies. These include a utility cloud computing environment (Cloudera) and analytic cloud components for big data (Accumulo, MapReduce, and Storm) [19, 20, 21, 22]. The knowledge stores themselves have many different representations conducive to varied types of data, including time-based stores, linked graphs, structured data stores, geospatial stores, metadata, and

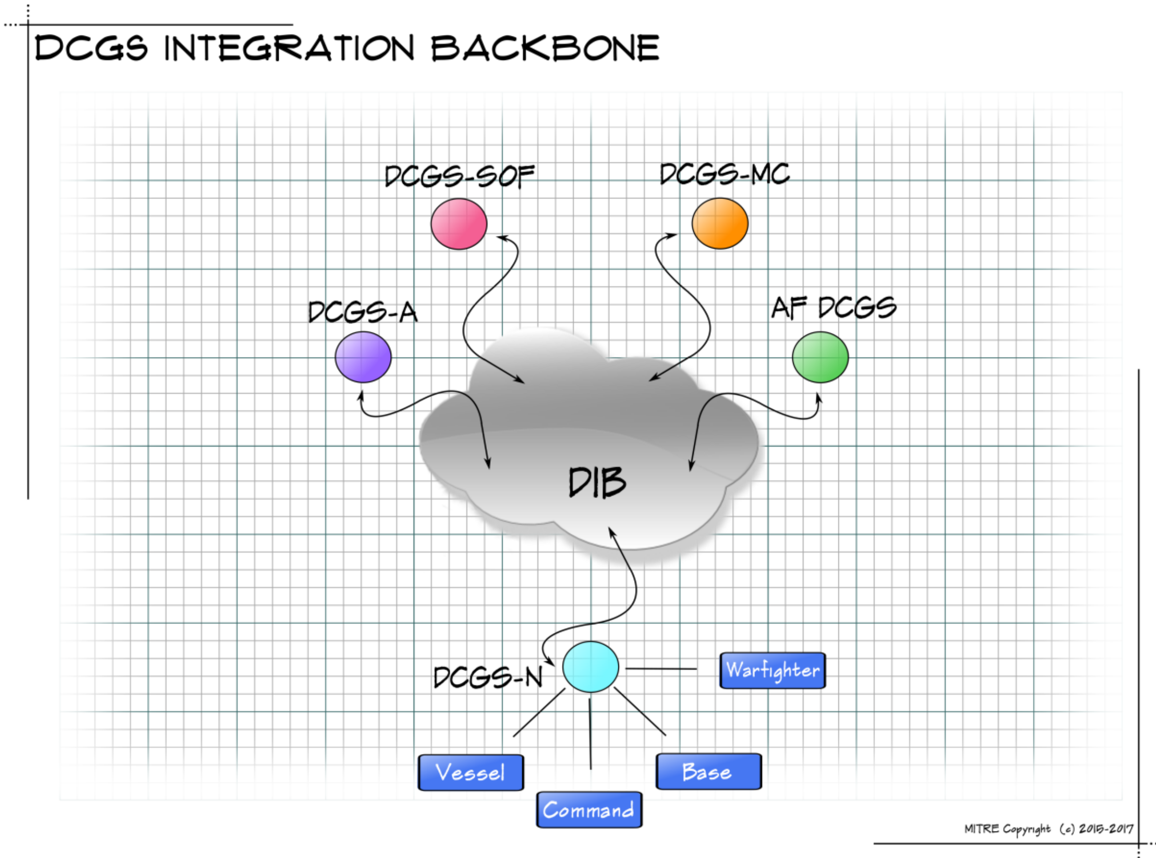

Figure 1: DCGS Integration Backbone (DIB) which bridges different DCGS weapon systems

provenance [21]. Exercises to assess TCRI happened this year as part of the Naval Integrated Tactical-Cloud Reference for Operational Superiority (NITROS) demonstration, and the capabilities may be fielded by 2019 or 2020 [23, 24, 25].

Another significant DCGS example is for the special operations forces (DCGS-SOF). As with other DCGS instances, the overarching goal is to connect intelligence data throughout the levels of SOF, from the warfighter to the leadership [26]. DCGS-SOF arises from a fusion of government and commercial solutions, for example MarkLogic, MetaCarta, and NetOwl, the last two performing data enrichment via content analysis and information extraction [27].

There is a final example of DCGS for the Marine Corps. DCGS-MC represents a collection of systems that provides the capability to process, exploit, analyze, and store data and intelligence products [28] across a portfolio that includes multiple families of intelligence capabilities (e.g., geospatial, all-source, signals) [29]. Also part of the composition of DCGS-MC is the Enterprise DIB Service (EDS), which serves as the interface to the DIB, allowing data and products to be stored and exposed within and beyond the enterprise [30].

Object Based Production and Activity Based Intelligence

The concepts of Object Based Production (OBP) and Activity Based Intelligence (ABI) run parallel to many of the intelligence and defense efforts described above. These emerged from the dual problems of intelligence often being organized or limited to its producer and the increasingly vast amount of data that must be digested to produce intelligence. The basic idea behind OBP [31] is to reorganize how intelligence is created to be centered on objects of interest and to disseminate based on this organization of intelligence. This approach expects different intelligence providers and owners to synthesize these objects collectively, so analysts can more readily acquire what they want to know, rather than requiring analysts to go to different sources individually. ABI [32, 31] is supplementary to OBP, relying on finding linkages between important entities and events across copious data from multiple intelligence sources to assemble the information that fits the OBP model and fill the gaps that might be incurred when only using a single source. Like many of the enterprise solutions described above, this pattern also employs underlying cloud technology to accommodate the volumes and high velocity of data and leans on the need for automated analytic tools to process this data into useful, enriched entities. In the Intelligence Community, the ongoing effort to provide OBP across the enterprise is called QUELLFIRE, which operates in a cloud environment (specifically Ghost Machine nodes) and employs Ozone Widget Framework (OWF) to create its objects before presenting analysts with visualizations assembled from intelligence derived from varied sources [33].

Centers for Medicare & Medicaid Services

While defense and intelligence agencies contend with copious data, there are also civilian federal agencies with programs or efforts that have access to and make use of big data sets. One representative example is the Centers for Medicare & Medicaid Services (CMS). To give a sense of scale, CMS had a Fiscal Year 2015 budget of $984.5 billion [34], which was primarily disbursed as Medicare, Medicaid, and CHIP medical claims reimbursement and benefits payments. Furthermore, the scale of data at CMS is quite large, where 4.5 million claims are processed daily [35]. CMS has developed some enterprise-wide capabilities to link and harmonize data, such as the Integrated Data Repository (IDR) [36] and Chronic Condition Warehouse (CCW) [37], but there are some notable challenges to these capabilities. While CMS provides enclaves [38], laboratories in which it allows researchers to combined its anonymized data with external data, there is potential for improvement regarding governance to ensure consistent use of the data.

At CMS, the Affordable Care Act (ACA) and its mandate to collect and integrate data from state exchanges [39] has intensified the data integrity challenge. Because of the nature of the data, there are also regulatory barriers to allowing wider access and data fusion. For instance, the Health Insurance Portability and Accountability Act (HIPAA) [40, 41] sets policies for protecting personally identifiable health information that, for instance, prevent sharing of data outside its original intent. Furthermore, CMS is driven by a legally required 30-day timeframe for processing claims [42], which can be at odds with its mission focus on access to better care, lowering tolerance for delays and obstacles to claims payments, such as fraud detection false positives. Still, there are ongoing and successful efforts in predictive analytics at CMS to catch fraud, saving billions in the past few years [43]. In addition to these policy and legal restrictions, there is an added technical challenge of analytic expediency and integrity, where automated decisions not only must be made, but understood, justified, and be verifiable [44].

Department of Veterans Affairs

One example of a non-military or intelligence agency effort that looks toward a wide-scale analytic architecture was encountered in discussions about the Department of Veterans Affairs (VA). There is an ongoing effort supporting this agency that involves building an ontology to influence future medical record schemas toward the goal of demonstrating the value of reasoning in a health care context [45]. This requires organizing current and future electronic medical records into this model for clinical use cases (as opposed to use cases such as fraud identification or billing). Further, the VA is also promoting an open source community called Open Source Electronic Health Record Alliance (OSEHRA) [46], which will help give third parties a baseline for interacting.

Elements of an Analytic Architecture

The above examination of a sample of analytic solutions across government reveals that, while there is in some agencies a desire for an enterprise-wide analytic architecture, there has yet to emerge established and ubiquitous solutions, which means it is an opportune time to consider and introduce alternatives to emerging solutions, bearing in mind what the survey reveals. In our case, we consider architectures that foster collaboration not only across government enterprise, but with the capacity for industry contribution. In particular, we seek to provide a hub for this collaboration, which we call a *Common Exchange Service*. This suggests three principal goals:

1. Providing a Common Exchange Service architecture to publish and subscribe to analytic results
2. Creating a public analysis exchange model that serves as a lingua franca for analytic results
3. Fostering the adaptation of different native analytic results into and out of the analysis exchange model

The Common Exchange Service architecture's principal components include 1) a *Pub/Sub Service*, which allows the publication of and subscription to analytic results either as part of workflows or in an ad hoc manner, 2) a *Transfer Service*, which is responsible for the bulk or streaming transfer of results between the Common Exchange Service and an external provider or consumer, 3) an *Adapter Service*, which converts results between a variety of native analytic models and the analysis exchange model and allows for usability across varied independently developed and executing analytic systems, and 4) a *Pub/Sub Store*, which temporarily holds results to be transferred. Eventually, this store will be supplemented with a persistent *Knowledge Store*, which may retain the most significant results that are required in perpetuity. Figure 2 depicts the layout of these components and their relationship with analytic services.

One advantage of this kind of exchange service is that individual analytics and workflows can operate independently from the overall enterprise, only accessing it intermittently as needed. Many real-world use cases suggest that handling this requirement is essential when you cannot guarantee a consistent connection to the enterprise from all access points at all times. Furthermore, analytics and workflows can be restricted to interact only when they have the permission and need for results outside their own environment. This allows both industry and government to create flexible relationships with multiple partners where data is shared when mutual interests are met or business arrangements have been established.

Analytic Execution

While the Common Exchange Service does not natively support automated analytics, it can be a part of analytic threads that make use of it as part of their pipelines or workflows, either publishing or consuming the analytic results it holds. A previous MITRE proposal for an architectural model for research, BlueRidge [47], described different patterns for handling and operating on data, knowledge, and results, and these appear relevant in this context too.

Following an assertion in the BlueRidge work, results once committed to the Pub/Sub Store should be treated as immutable, which is to say that while they can be removed or accessed, they cannot be altered; analytically altered results would in turn become a new result set. While publish and submit are expected to be the main actions that center around

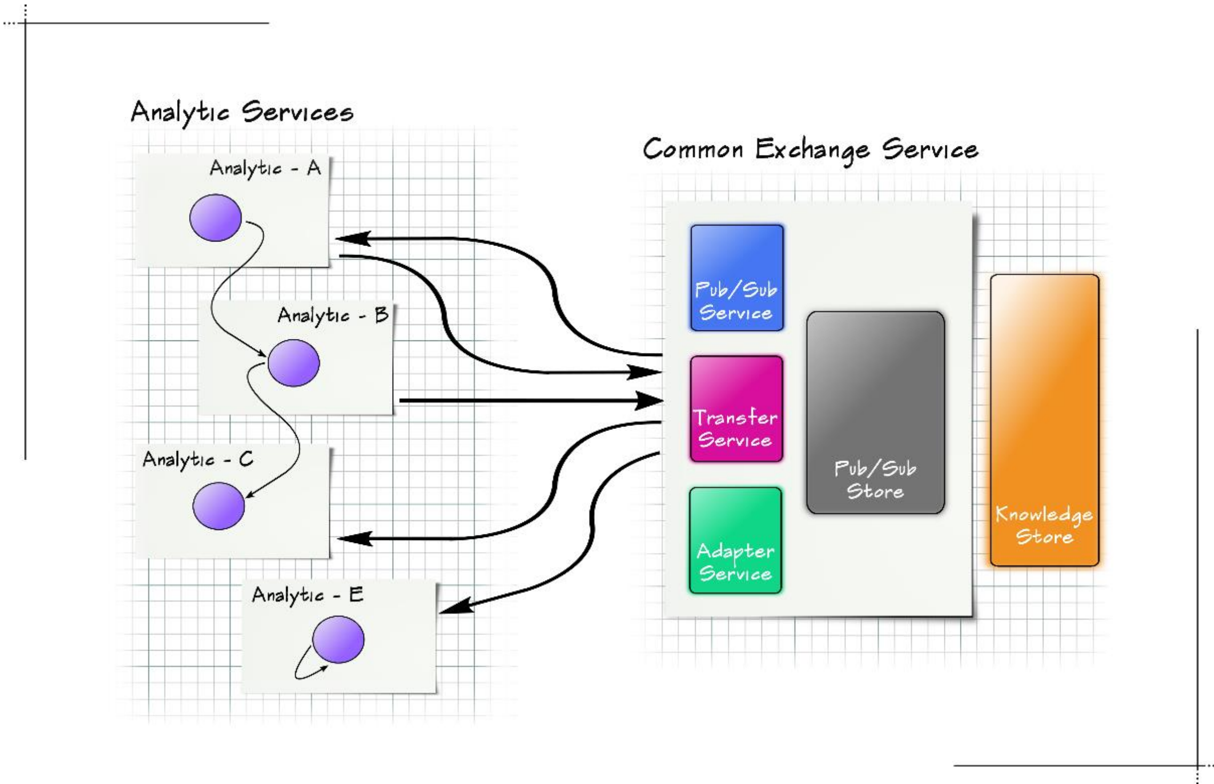

Figure 2: Common Exchange Service and its principal components

this hub, it is possible, as in the BlueRidge work, to support more advanced manipulation and processing of results, with flows of activity producing—and potentially reusing—results in a sequence of analytics (pipeline flow), aggregating or merging results into new results (aggregation flow), or changing result objects into something different (alteration flow). Figure 3 (from the BlueRidge documentation) shows these different processes, and helps to delineate where data (or stored results in this case) are considered immutable (in the data bus which is analogous to the Pub/Sub Store) versus not. Naturally, access to the Pub/Sub Store may also be simpler and ad hoc.

Exchange Models for Analytic Results

An element that recurs throughout the enterprises where analytic architectures are being explored is the need for a common representation and consistent definition of knowledge objects that make up results. While the Common Exchange Service could store results in their raw output form, an analysis exchange model and an evolving library of

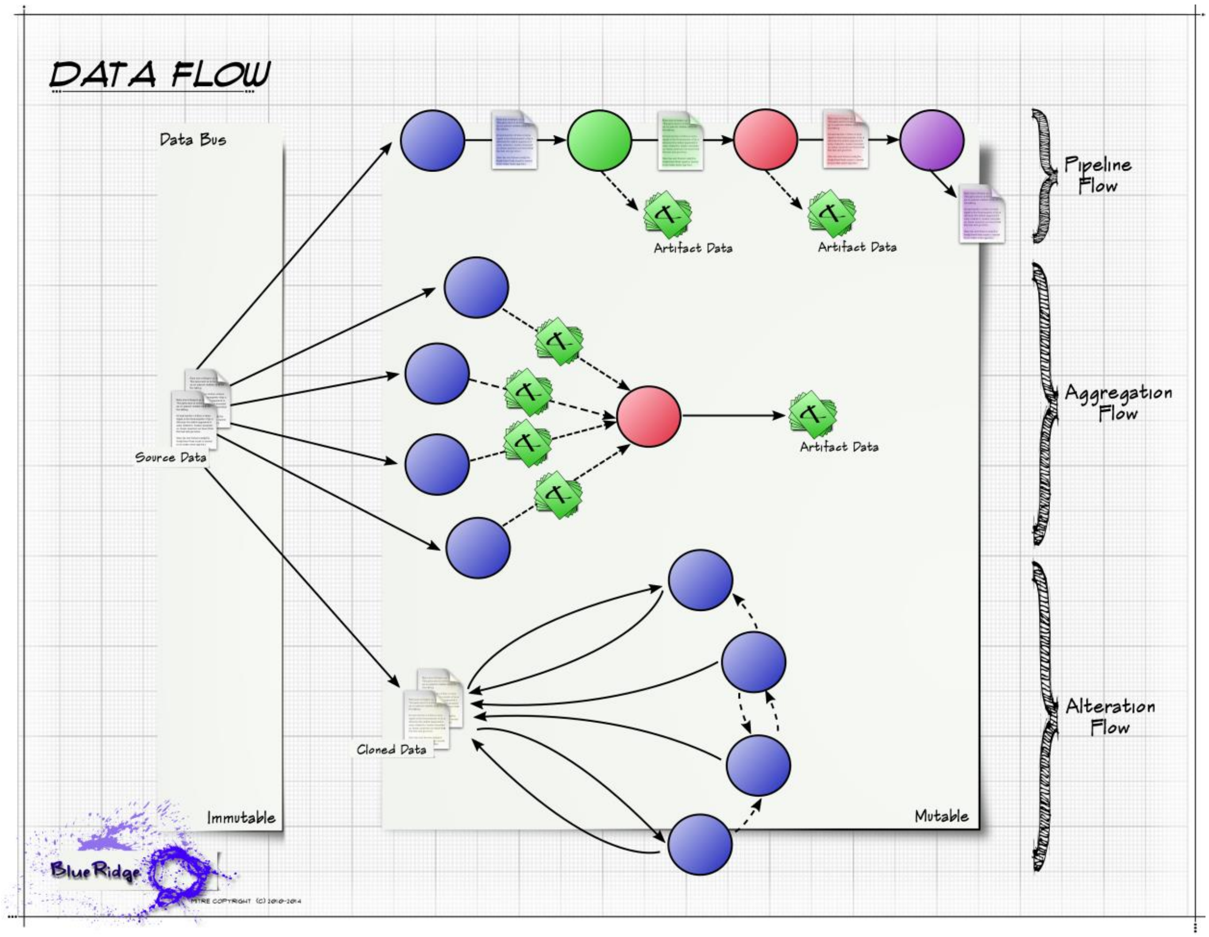

Figure 3: Flows of data (or results) as originally presented in MITRE BlueRidge work [48]

adaptation software that keeps pace with changing native models means useful interactions can spin up more quickly and without the burden of adaptation necessarily having to fall to one partner or the other. Adaptation software can be provided by companies should they wish to retain a close hold on proprietary processes and models or they can be engineered by outside parties if the issue is one of aligning a published standard with the analysis exchange model.

While it would be ideal to use a single comprehensive ontology that could draw on and reason over results across varied domains for any conceivable use case, to our knowledge no such ontology exists. Rather than making the attempt to design such an ontology or import an existing ontology to serve in this capacity, it is better to acknowledge that we will begin with a narrower use case and develop (or import) a data model and ontology that fits the parameters of the kind of results that are likely to be available from the different analytics and desirable for the use case. Depending on the scope, existing ontologies with broad coverage, such as Cyc [48, 49], SUMO [50], or Basic Formal Ontology (BFO) [51], might serve as an upper ontology, but it is quite likely that a tailored, more domain specific ontology will be better suited for a specific use case.

Storage

Another element of the Common Exchange Service where the choices are largely driven by the use cases is choosing the proper solution for storage. The underlying model is quite malleable, and scalable storage solutions (for example, Accumulo [52], which is employed in several developing architectures) can be applied, or if the result sets are smaller, less cumbersome and more targeted solutions can be implemented. Again, while there is an implicit assumption the results will consist of knowledge objects, these may be encoded in storage architectures such as object stores, key-value stores, graph stores, or document stores.

Other issues in storage that are dictated by the use case include the volatility of the results (i.e., how long does the exchange preserve the results?), the rate of ingest, and the manner of ingest, whether results are uploaded in bulk or following a streaming model. While the Common Exchange Service's basic design should accommodate different activities, these activities will determine what solutions are implemented for the Transfer Service, the Pub/Sub Store (where results are temporarily held) and the Knowledge Store (where significant results are held long-term).

Conclusions

This paper has explored the current state of analytic architectures across various government agencies, each with its own needs and objectives for converting different scales of data into valuable knowledge or intelligence. There are no settled solutions for any agency, and the time is right to offer alternatives that foster collaboration between government agencies and different industry partners that can provide essential capabilities, relationships that can be mutually beneficial to all parties involved. Our initial proposal is to consider the development of a reference implementation of an analytic architecture supported by the Common Exchange Service, which can serve as a hub for this collaboration and can include contributions from across industry while remaining flexible enough to suit several different use cases and specific implementations of technology to support its model.

Acknowledgements

This paper reflects insights provided by several individuals at MITRE. In particular, we would like to thank our colleagues who spoke at length with us on the different topics, including Kathryn Aitkenhead, L. Karl Branting, Dick Dramstad, Brad Evans, Mike Fleckenstein, Richard Games, Gary Glessner, Ken Griesi, Armand Kinslow, David Lee, Beth Masimore, Leo Obrst, Julia Packer, John Prange, Ted Sienknecht, Paul Silvey, Mark Simons, Merwyn Taylor, Naimish Thakkar, Andy Thompson, and Lisa Tompkins-Brown.

References

[1] K. Quinn, "A Global Intelligence Enterprise," Trajectory Magazine, 2016. [Online]. Available: http://trajectorymagazine.com/defense-intelligence/item/2116-a-global-intelligence-enterprise.html. [Accessed 14 6 2016].

[2] A. Corrin, "DI2E Framework Seeks to Unite a Wealth of Intelligence Data," GCN, 6 May 2011. [Online]. Available: https://gcn.com/Articles/2011/05/03/Defense-IT-1-intelligence-enterprise.aspx. [Accessed 15 6 2016].

[3] INSA, "IC ITE: Doing in Common What is Commonly Done," 2 2013. [Online]. Available: http://www.ctolabs.com/wp-content/uploads/2014/01/ICITE_Doing_In_Common-1.pdf. [Accessed 15 6 2016].

[4] P. Tucker and F. Konkel, "Obama Pushes for Greater Intel Sharing in New Strategy," 9 2 2015. [Online]. Available: http://www.defenseone.com/technology/2015/02/obama-pushes-greater-intel-sharing-new-strategy/104877/. [Accessed 2 11 2016].

[5] K. Quinn, "IC ITE Moves Forward," 9 10 2013. [Online]. Available: http://trajectorymagazine.com/got-geoint/item/1570-ic-ite-moves-forward.html. [Accessed 2 11 2016].

[6] Office of the Under Secretary of Defense for Intelligence, "Defense Intelligence Information Enterprise (DI2E): A "Bridge" between JIE and IC ITE," 22 5 2014. [Online]. Available: http://www.afei.org/PE/4A07/Documents/DI2EPlugfest2014-JimMartinBrief-v0.pdf. [Accessed 5 11 2016].

[7] Office of the Under Secretary of Defense for Intelligence, "DI2E Framework Standard Brief," 10 5 2013. [Online]. Available: http://c4i.gmu.edu/eventsInfo/reviews/2013/pdfs/AFCEA2013-Lane.pdf. [Accessed 5 11 2016].

[8] U.S. Air Force, "Air Force Distributed Common Ground System," 13 October 2015. [Online]. Available: http://www.af.mil/AboutUs/FactSheets/Display/tabid/224/Article/104525/air-force-distributed-common-ground-system.aspx. [Accessed 8 6 2016].

[9] "Air Force Distributed Common Ground System (AF DCGS)," 2014. [Online]. Available: http://www.dote.osd.mil/pub/reports/fy2014/pdf/af/2014af_DCGS.pdf. [Accessed 3 11 2016].

[10] P. Welsh, "Open architecture bringing benefits to AF DCGS," 23 8 2016. [Online]. Available: http://www.hanscom.af.mil/News/Article-Display/Article/922561/open-architecture-bringing-benefits-to-af-dcgs. [Accessed 6 11 2016].

[11] U.S. Air Force, "Air Force ISR 2023: Delivering Decision Advantage," 2013. [Online]. Available: http://www.defenseinnovationmarketplace.mil/resources/AF-ISR_2023.pdf. [Accessed 3 8 2016].

[12] esri, "The Synergy of Networkcentric Technologies, GIS, and the Distributed Common Ground System's Integration Backbone (AF DCGS Block 10.2)," [Online]. Available: http://www.esri.com/~/media/Files/Pdfs/industries/defense/pdfs/air_force_dcgs.pdf. [Accessed 9 6 2016].

[13] T. Fitzgerald, "Distributed Common Ground System - Army Supporting Air Force and Army Interoperability," 29 6 2015. [Online]. Available: https://www.army.mil/article/151379/Distributed_Common_Ground_System_ ___Army_Supporting_Air_Force_and_Army_Interoperability/. [Accessed 10 6 2016].

[14] U.S. Army, "Distributed Common Ground System - Army," 16 1 2014. [Online]. Available: https://www.army.mil/standto/archive_2014-01-16/. [Accessed 10 6 2016].

[15] J. Judson, "Army Releases RFP for DCGS-A Increment 2," 23 12 2015. [Online]. Available: http://www.defensenews.com/story/defense/land/army/2015/12/23/army-releases-rfp-dcgs--increment-2/77843320/. [Accessed 10 6 2016].

[16] Booz Allen Hamilton, "People Purpose Passion: Impact Report 2016," 2016. [Online]. Available: https://www.boozallen.com/content/dam/boozallen/impact16/Booz-Allen-Hamilton-Impact-Report-2016.pdf. [Accessed 2 11 2016].

[17] Booz Allen Hamilton, "Lights On," 27 10 2015. [Online]. Available: http://www.boozallen.com/stories/lights-on. [Accessed 2 11 2016].

[18] "Tactical Cloud Reference Implementation," [Online]. Available: https://www.sbir.gov/taxonomy/term/156063. [Accessed 10 6 2016].

[19] B. Brewin, "The Navy Wants a Tactical Cloud," Defense One, 25 9 2014. [Online]. Available: http://www.defenseone.com/technology/2014/09/navy-wants-tactical-cloud/95129/. [Accessed 3 8 2016].

[20] ONR, "ONR BAA Announcement # ONRBAA14-011," 2014. [Online]. Available: http://www.onr.navy.mil/~/media/Files/Funding-Announcements/BAA/2014/14-011.ashx. [Accessed 2 11 2016].

[21] ONR, "Data Focused Naval Tactical Cloud," 24 6 2014. [Online]. Available: http://www.onr.navy.mil/~/media/Files/Funding-Announcements/BAA/2014/14-011-Attachment-0001.ashx. [Accessed 3 11 2016].

[22] K. McCaney, "Navy wants to take big data into battle," 24 6 2014. [Online]. Available: https://defensesystems.com/articles/2014/06/24/navy-onr-big-data-ecosystem.aspx. [Accessed 3 11 2016].

[23] R. Boland, "NITROS Boosts Information Warfare at Trident Warrior," 18 7 2016. [Online]. Available: http://www.navy.mil/submit/display.asp?story_id=95744. [Accessed 3 11 2016].

[24] "US Navy tests NITROS information warfare system at Trident Warrior," 19 7 2016. [Online]. Available: http://navaltoday.com/2016/07/19/us-navy-tests-nitros-information-warfare-system-at-trident-warrior/. [Accessed 4 11 2016].

[25] S. Heller and C. Miller, "Space and Naval Warfare Systems Center Atlantic Update," 7 9 2016. [Online]. Available: http://www.public.navy.mil/spawar/Atlantic/Documents/Industry/CO%20ED_%20SBIOI%209-8-16.pdf. [Accessed 4 11 2016].

[26] "Special Operations Forces Industry Conference: Special Reconnaissance, Surveillance, and Exploitation," [Online]. Available: http://www.defenseinnovationmarketplace.mil/resources/SOCOM_Intelligence.pdf. [Accessed 3 11 2016].

[27] "All-Source Analytical Environment (ASAE) Request for Proposal," 15 10 2015. [Online]. Available: https://www.fbo.gov/?id=2028b593ec54a62f877678b9c0aae3f6. [Accessed 3 11 2016].

[28] "Distributed Common Ground-Surface System Marine Corps (DCGS-MC)," US Marine Corps, 4 6 2015. [Online]. Available: https://marinecorpsconceptsandprograms.com/programs/intelligence-surveillance-and-reconnaissance/distributed-common-ground-surface-system. [Accessed 10 6 2016].

[29] S. Adams, "Planning Brief to Industry," 22 9 2015. [Online]. Available: http://www.marcorsyscom.marines.mil/Portals/105/CorpComDirectorate/Command%20 Briefs/112-%20PM%20MI.pdf. [Accessed 5 8 2016].

[30] "Distributed Common Ground System - Marine Corps (DCGS-MC)," 2014. [Online]. Available: http://www.dote.osd.mil/pub/reports/fy2014/pdf/navy/2014dcgs-mc.pdf. [Accessed 5 8 2016].

[31] P. Buxbaum, "Organizing the "Knowns"," 2 2 2015. [Online]. Available: http://www.kmimediagroup.com/gif/424-articles-gif/organizing-the-knowns/6361-organizing-the-knowns. [Accessed 6 11 2016].

[32] C. P. Atwood, "Activity-Based Intelligence: Revolutionizing Military Intelligence Analysis," National Defense University Press, Joint Force Quarterly 77, 4 2015. [Online]. Available:

http://ndupress.ndu.edu/Media/News/NewsArticleView/tabid/7849/Article/581866/jfq-77-activity-based-intelligence-revolutionizing-military-intelligence-analys.aspx. [Accessed 23 6 2016].

[33] C. Johnston, "Modernizing Defense Intelligence: Object Based Production and Activity Based Intelligence," 27 6 2013. [Online]. Available: https://www.ncsi.com/diaid/2013/presentations/johnston.pdf. [Accessed 23 6 2016].

[34] CMS, "CMS Fast Facts July 2016 Version," [Online]. Available: https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/CMS-Fast-Facts/index.html. [Accessed 03 08 2016].

[35] CMS, "Healthcare Integrated General Ledger Accounting System (HIGLAS)," [Online]. Available: https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/HIGLAS/index.html?redirect=/higlas/. [Accessed 3 8 2016].

[36] "CMS Integrated Data Repository (IDR)," 17 4 2013. [Online]. Available: https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/IDR/index.html. [Accessed 4 11 2016].

[37] "Chronic Conditions Data Warehouse," [Online]. Available: https://www.ccwdata.org/web/guest/condition-categories. [Accessed 4 11 2016].

[38] E. C. Hair, "Accessing CMS Claims Records: Data Enclave as a Virtual RDC," 2011. [Online]. Available: http://www.ispor.org/meetings/baltimore0511/presentations/WS5_Hair.pdf. [Accessed 5 11 2016].

[39] "State Data Collection and Reporting Requirements in the Affordable Care Act," 20 4 2012. [Online]. Available: http://www.shadac.org/sites/default/files/Old_files/ACA_Reporting_Data_Collection_Requirements_Draft_4_20_12_meeting.pdf. [Accessed 5 11 2016].

[40] 104th Congress Public Law 191, "Health Insurance Portability and Accountability Act of 1996," 1996. [Online]. Available: https://www.gpo.gov/fdsys/pkg/PLAW-104publ191/html/PLAW-104publ191.htm. [Accessed 24 6 2016].

[41] CMS, "HIPAA Basics for Providers: Privacy, Security, and Breach Notification Rules," 8 2016. [Online]. Available: https://www.cms.gov/Outreach-and-Education/Medicare-Learning-Network-MLN/MLNProducts/Downloads/HIPAAPrivacyandSecurity.pdf. [Accessed 5 11 2016].

[42] CMS, "Medicare Claims Processing Manual: Chapter 1 - General Billing Requirements," 29 4 2016. [Online]. Available: https://www.cms.gov/Regulations-and-Guidance/Guidance/Manuals/Downloads/clm104c01.pdf. [Accessed 5 11 2016].

[43] S. Agrawal, "$42 Billion Saved in Medicare and Medicaid Primarily through Prevention," 20 7 2016. [Online]. Available: https://blog.cms.gov/2016/07/20/42-billion-saved-in-medicare-and-medicaid-primarily-through-prevention/. [Accessed 5 11 2016].

[44] Centers for Medicare & Medicaid Services, "Data Analytic Capabilities Assessment for Medicaid Program Integrity," 9 2014. [Online]. Available: https://www.cms.gov/medicare-medicaid-coordination/fraud-prevention/fraudabuseforprofs/downloads/data-analytic-assesstoolkit-092214.pdf. [Accessed 5 11 2016].

[45] S. Stoutenberg, L. Obrst, J. Nebeker, M. Ward, S. Bennett, L. Seligman, S. Sebastian and M. Jones, "Developing a Clinical Care Ontology for the Veterans Health Administration," in *Ontology in Practice: The Fourth Clinical and Translational Science Ontology Workshop*, Charleston, SC, 2015.

[46] "OSEHRA," [Online]. Available: https://www.osehra.org/. [Accessed 6 7 2016].

[47] J. Jubinski, R. Winder and A. M. O'Hanlon, "Architectural Model for MITRE Research: BlueRidge," 12 2015. [Online]. Available: https://www.mitre.org/sites/default/files/publications/pr-15-1254-architectural-model-mitre-research-blueridge.pdf.

[48] D. B. Lenat, "CYC: A Large-Scale Investment in Knowledge Infrastructure," *Communications of the ACM,* vol. 38, no. 11, pp. 33-38, 1995.

[49] C. Matuszek, J. Cabral, M. Witbrock and J. DeOliveira, "An Introduction to the Syntax and Content of Cyc," in *AAAI Spring Symposium: Formalizing and Compiling Background Knowledge and Its Application to Knowledge Representation and Question Answering*, 2006.

[50] I. Niles and A. Pease, "Towards a Standard Upper Ontology," in *Proceedings of the International Conference on Formal Ontology in Information Systems*, 2001.

[51] R. Arp, B. Smith and A. D. Spear, Building Ontologies with Basic Formal Ontology, The MIT Press, 2015.

[52] "Apache Accumulo," [Online]. Available: http://accumulo.apache.org/. [Accessed 10 6 2016].

Appendix

Government Survey

The following is the survey that was provided to MITRE staff supporting different government organizations. This served as a starting point for discussions for collecting information.

Architecture Survey and Review is one of several formal studies currently underway by MITRE's ***Analysis Technology Industry Roundtable***. This survey is being championed by MITRE (Ransom Winder, Joseph Jubinski, Angela O'Hanlon).

Our team is gathering information on analytic architectures used across government sponsors. As part of this study, we are surveying analytic architectures, and one element of our task is to include what we can discover about the practice in government spaces.

The questions below are intended to capture information about your project and analytic architecture. Our interview and discussion will capture much more detail, as we realize that each architecture, system, and environment differs.

- In your government sponsor, are there organization-wide analytic architectures that developers are encouraged, expected, or mandated to use?
- If there are, how long have these architectures been in place and how apt are they to change?
- If there are not, are there organization-wide analytic architectures being proposed? Or are there solutions that are in practice used by a majority (or at least a significant number) of developers?
- What are your sponsors looking for in a solution such as this? Are there any gaps that an analytic architecture can fill for your government sponsor?
- Do you have an OV-1 that we can use to illustrate your architecture, particularly one that is publicly releasable?
- Do you recommend additional MITRE or Government POCs from your organization to discuss this arch?
- Do you recommend additional architectures for this study?