The Analysis Exchange Technology

A Design and Solution for Software Interoperability by the

Technology & Innovation Roundtable:

Selected Pilot Work with the US Army and US Air Force

November, 2019

Approved for Public Release; Distribution Unlimited. Public Release Case Number 19-2795.

MTR190652

Authors

Mr. Brian Biesecker, Esri
Mr. Charles Brown, IBM
Mr. Jeffrey Eggers, DISL, US Air Force A2/6
Dr. Adam Etches, IBM
Ms. Beth Flaherty, SYSTRAN
Mr. Joseph Jubinski, MITRE
Dr. William Niehaus, NetOwl
Dr. Angela O'Hanlon, MITRE
Colonel Bradley Readnour, US Air Force A2/6
Mr. Gary Retzlaff, US Army TRADOC G2
Mr. John Stultz, SAS
Dr. Ransom Winder, MITRE

The Analysis Exchange technology design, creation, and issue was a collaborative effort over the course of three years, including multiple authors. Our authors are listed in alphabetical order and considered mutual contributors to the overall design, success, and research effort of the Analysis Exchange technology.

Government sponsors who supported and gave guidance on use cases for the Analysis Exchange pilots included Col. (Ret.) Jeffrey Eggers (DISL, USAF A2/6), Col. Bradley Readnour (USAF A2/6), Mr. Gary Retzlaff (CIM, Army TRADOC G2), and Ms. Kaye Darone (Army TRADOC G2).

Industry partners and Roundtable members who contributed to the studies and pilots related to the Analysis Exchange included Mr. Charlie Brown (IBM), Mr. John Stultz (SAS), Dr. Adam Etches (IBM), Dr. William Niehaus (NetOwl), Mr. Brian Biesecker (Esri), Ms. Jennifer Flather (Basis), Mr. Ed Kenschaft (Basis), Mr. Lorne Hanson (Centrifuge), Mr. Steve Panzer (Centrifuge), Mr. Joon Kim (Cloudera), Mr. Joel Valverde (Cloudera), Mr. David Waldrop (ICG Lux), Mr. Scott Graff (Microsoft), Mr. Mike Mullins (Microsoft), Mr. Matthew Chandler (Palantir), Mr. Bryant Choung (Palantir), Mr. Michael Ramirez (Recorded Future), Mr. C.W. Walker (Recorded Future), Mr. Brian Adams (SAP NS2), Mr. Brian Major (SAP NS2), Mr. Frank McGuinness (SAP NS2), Mr. Nicholas Riccio (SAP NS2), Mr. Bill Stewart (SAP NS2), Mr. Kelly Hu (SitScape), Mr. Joe Marinich (SitScape), Mr. Satish Abburi (System Soft Technologies), Mr. Kranthi Gajjala (System Soft Technologies), Mr. Eric Warner (System Soft Technologies), Mr. Hugh Aitchison (SYSTRAN), Ms. Beth Flaherty (SYSTRAN), Mr. Juan Johnson (Tableau), Mr. Gene McCluskey (Tableau), Mr. Ross Paulson (Tableau), Mr. Blake Ramsey (Tableau), Mr. David Sears (Tableau), Mr. Andy Russell (Thomson Reuters), Ms. Dawn Scalici (Thomson Reuters), and Mr. Derek Smith (Thomson Reuters).

MITRE architects and designers dedicated to the Analysis Exchange development were Dr. Angela O’Hanlon (Roundtable Director and Chair), Mr. Joseph Jubinski, Dr. Ransom Winder, Mr. Akash Trivedi, Mr. Elmer Armijo Rivera, and Mr. Marc Ubaldino.

This document was written and prepared by Dr. Ransom Winder, Dr. Angela O’Hanlon, Mr. Joseph Jubinski, Col. Bradley Readnour, Mr. Gary Retzlaff, Mr. John Stultz, Ms. Beth Flaherty, and Dr. William Niehaus.

This page intentionally left blank.

Table of Contents

Introduction

Bottom Line Up Front – Why Should I Care?

Every government agency relies on data analytics for good decision making and uses analysis and analytic technologies from a wide variety of technology companies. However, putting together cohesive systems from multiple analytic tools can be a challenge, especially when it comes to meeting the increasing demands for agility and rapid outcomes. The Technology & Innovation Roundtable[1] is a consortium of industry partners committed to addressing government priorities, including facing the above challenge.

The Roundtable's solution is the Analysis Exchange, which works toward achieving the following benefits: 1) using a more interoperable, consistent, and repeatable model to allow vendors to compete for acquisitions more efficiently, saving time and money, 2) facilitating products and systems to become more effective at meeting government needs and requirements with less investment of resources, 3) breaking down data silos that multiple analytic tools typically create, 4) encouraging the quick adoption of consistent standards for the analytic community, saving both the government and industry time and money, and 5) creating enriched products for the government based on the best solutions provided by different companies.

The Analysis Exchange work has evolved from addressing use cases that demonstrate its collaborative capability in a variety of domains to supporting an FY19 pilot for the United States Army TRADOC, addressing its mission workflows in a cloud environment, as open source data undergoes analysis and enrichment and ultimately visualization with capabilities provided by different industry partners.

Contents

This paper first describes the motivations that lie behind this challenge from the government and industry perspective. It then summarizes the insights from the Roundtable's initial studies into the landscape of

[1] www.mitre.org/roundtable, formerly the Analytic Technology Industry Roundtable

architecture and analytic tools. The paper then highlights the technology underlying its solution to the challenge: the Analysis Exchange. A description of the use cases is then explained. A section describing the ongoing pilot effort in FY19 is provided as well. The paper then concludes with a discussion of metrics of success and the significance of the Analysis Exchange's impact to government and industry.

Motivation

Both government and industry are motivated to address longstanding and emerging issues with interoperability between analytic capabilities in a landscape where the need to achieve quick results covering a broad spectrum of analytic capability continues to accelerate. This section explains what drives both providers and customers in this environment.

Government Incentives

Government agencies have established mandates for being cost effective and efficient. They do not want to be left behind as technology improves and new features are available at lower prices on analytic solutions. Given the rigors of government acquisition of these solutions, flexibility is sacrificed. This accentuates the challenge when acquisition of multiple solutions from different vendors is required to solve the government's problems, especially when technical incompatibilities exist between what the government uses today, wants to use, and could use. While the government could create a custom solution and build their own tools within that solution, this requires significant effort with high costs and long times to delivery. This is infeasible for many agencies that have either limited budgets or mission goals that shift. Often that solution is proprietary and is behind state-of-the-art when delivered. Therefore, finding a way to overcome the challenges to leverage commercial solutions is desirable.

The above suggests that, from the government perspective, shortfalls in industry interoperability exist. Vendor analytic solutions often include a unique data model and single reference database. This creates a stovepipe effect for data where the data is not easily transported to other analytic tools. Vendor analytic solutions are good at solving niche analytic problems, and therefore the government is required to make tough trade decisions on their analytic tool investment. For example, tool X handles 60% of a government requirement and tool Y handles 70% with some overlap in capability. Government is forced to down-select to a single tool with a known requirement gap because interoperability of data does not exist to make investment in both tools a viable solution. The rate of change of tools and how they are applied also affects government decisions. Emerging requirements and technology advancements create a significant changeover cost decision for the government. Without analytic tool interoperability, the government will often stay with a legacy analytic tool rather than incur the full cost of switching vendors.

Consider a more specific case in point. Post 9/11, the emphasis on interoperability was set aside to more quickly deliver solutions that might save American lives on the ground. This led to a plethora of new sensors being adopted by the Services, generally without concern if they would interoperate with what already existed. An example of this is in the unmanned aircraft area. While the most famous of these systems was the Predator, many smaller systems were adopted and deployed, and nearly every one of them was proprietary in some regard. Even though the military wanted to adopt a single control station to be able to fly multiple types of unmanned aircraft, due to proprietary differences in the systems, this was impossible. Equally problematic, the data produced by most of these new sensors was in a proprietary format and could not be fused with similar data. This led to standalone analytical stations in the middle of intelligence floors trying to perform multi-INT analysis.

Full analytic interoperability would have a significant impact on government organizations. Data would be easily shared between analytic tools within a workflow. The government could fulfill 100% of their analytic tool requirement without making tough trade decisions. Finally, the government could take advantage of technology advances with analytic tools and meet emerging requirements in an agile fashion.

The capacity for full interoperability opens new possibilities for analysis. In United States Air Force (USAF) intelligence, surveillance, and reconnaissance (ISR), one of the biggest challenges is being overwhelmed with data. New sources of data are constantly being provided and the USAF struggles to find ways to turn that data into decision quality information by including it with data derived from countless other sources. Machine-to-machine correlation of data would go a long way towards allowing humans to be analysts, rather than using them to piece together data from multiple systems. A typical analyst may only spend 10% of his or her time doing the hard analytical work humans do better than machines, and 90% of their time finding and consolidating the data to be able to do analysis. In the aspirational state, this ratio would be reversed.

Underlying technological requirements vary for each government organization. As a case in point US Army TRADOC has specific needs for tasks involving considerable data that requires analysis. Analytic tools, both commercial and government, and the underlying analytic platform must be cloud capable. As the government is transitioning to a cloud (cloud hybrid) architecture, this is crucial. Standard industry ingest rates must be met and storage must be on the order of tens of terabytes (TBs). Execution time must be within minutes, producing results that are published to data repositories and become available to analysts through customizable visualization tools.

With the above requirements in mind, there are many analytic capabilities that the government desires, spanning the entire ingest-to-visualization workflow. These include natural language processing for text extraction and meta-data creation from unstructured data, structured data parsing and classification, data discovery, data analytics that include both unstructured (word use – modeling) and structured data (statistical analysis), and the ability to produce graphs, charts, timelines, maps, and link diagrams. The ability to display what analysts want to see at the end of the workflow and hide what does not interest them is also key to the end visualization.

Industry Motivation

The Roundtable's industry partners wish to meet the government's requirements in a timely fashion with the lowest cost for rework involved. This motivation drives them to seek a solution that can achieve certain benefits:

- Infrastructure interoperability and portability
- Acceleration of deployment
- Systems and user collaboration

The first desired benefit is infrastructure interoperability and portability, namely that what is developed as a solution can use existing components that provide the solution with minimal rework and that this solution itself can be applied again and again. The second desired benefit is an accelerated deployment of solutions, overcoming the tremendous costs often associated with acquisition, integration, and adoption. The third desired benefit is collaboration between systems and users, essentially defeating the tendency for solutions to exist in isolation or as part of fixed stovepipes.

The Roundtable's industry partners have many business justifications for working together in order to achieve these benefits. One business justification is the intention to "integrate and augment" and not to "rip and replace." In order to compete, industry must have strong partnerships and ways to integrate in order to be viable in today's markets. For analytic requirements focused on specific activities, it means smaller companies with targeted solutions can join with one another or with larger companies with more diverse offerings, thereby expanding the capabilities for the client.

This desire establishes requirements for how to interoperate, looking for solutions with low cost in money and time. Industry partners of the Roundtable seek a solution with high stability, reusability, and extensibility, which means it can be accredited once—or at least infrequently. Because analytic software itself often proceeds through versions rapidly, the industry partners prefer a solution that can itself be stable and accommodating of those changes without rework of an entire infrastructure using individual tools.

Nevertheless, there are challenges and concerns related to each of the different benefits the Roundtable and its partners seek. One of the challenges of "infrastructure interoperability" is the alignment of the business processes with the workflows of users and business/mission drivers and the software that is leveraged to meet the business/mission requirements. Even with the most interoperable environment where an open architecture is deployed, failures can occur if the business processes are not in sync.

Another challenge with achieving interoperability is the extent to which leadership is aligned with the users who are using the software. For example, when leadership is forcing usage of software that is incompatible with the business processes and or the skill sets of the users, users may leave due to their frustration or get creative and create "work arounds" to get their work done. These work around tactics often create more silos of disparate systems that inhibit the best of intentions that were originally part of the vision and executive buy in. Executives who can adapt to the reality of what end users need to get their

Challenges of infrastructure interoperability

- Alignment of the business process with workflows
- Leadership alignment with users
- Alignment with different providers of analytic solutions
- Coordinating vision within a team with different analytic capabilities

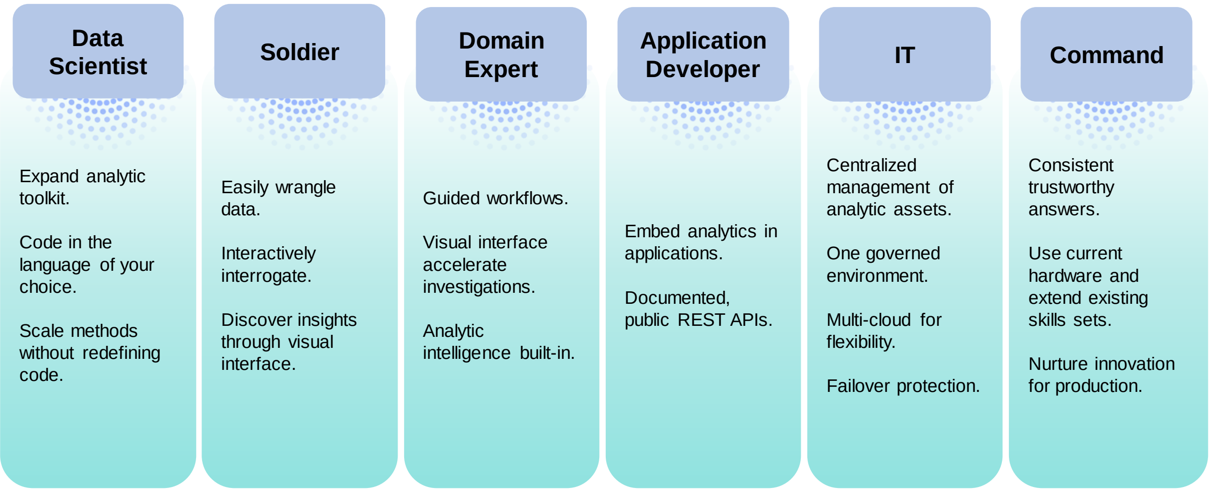
work done and pivot even if it goes against what they were sold and then championed “as the right move” help mitigate this type of end user behavior. This adaptability typically leads to interoperable processes or building systems that work. The challenge of aligning leadership, users, and business process also suggests another challenge, namely the importance of the customers openly and clearly expressing what they want to achieve so the right combination of solutions can be applied.

Interoperability challenges also exist between different providers of analytic solutions. Even with a clear goal provided by a customer, coordination within a team with different analytic capabilities requires a vision of how to design a workflow that will be consistently useful and align with the business process. In the case of more focused tools, this is especially important because with such a specific emphasis, their role in a larger system must be clear from the outset.

With respect to achieving an “acceleration of deployment,” an important lingering question is whether the current acquisition and accreditation practices can keep up with the rate of change that now occurs in commercial-off-the-shelf (COTS) analytic software. Traditionally slow acquisition and accreditation runs the risk of solutions being adopted after they are already surpassed by newer versions. This, of course, assumes there are “other direct costs” (ODC) funds to cover the costs of purchasing COTS analytic software. The challenge here is that, in the context of government customers, COTS tools end up competing with the development of government-off-the-shelf (GOTS) software or software derived from open-source, which incur costs of their own but are perceived differently.

Unrelated to challenges surrounding funds and acquisition, another challenge arises when considering a shared infrastructure for collaboration once it is deployed, namely who holds the responsibility for its installation, training users to work with it, and any long-term maintenance and refresh it requires.

Another set of challenges are encountered when considering the final benefit the Roundtable strives to achieve. One challenge related to achieving the benefit of “systems and user collaboration” is when an “integrated technology” infrastructure does not support current releases of industry standard platforms such as Java, Microsoft Component Object Model (COM), Distributed Component Object Model (DCOM), .NET, open API for Web integration, API support for the message oriented middleware (MOM) software, Lightweight Directory Access Protocol (LDAP) and Web Distributed Authoring and Versioning



Data Scientist

Expand analytic toolkit.

Code in the language of your choice.

Scale methods without redefining code.

Soldier

Easily wrangle data.

Interactively interrogate.

Discover insights through visual interface.

Domain Expert

Guided workflows.

Visual interface accelerate investigations.

Analytic intelligence built-in.

Application Developer

Embed analytics in applications.

Documented, public REST APIs.

IT

Centralized management of analytic assets.

One governed environment.

Multi-cloud for flexibility.

Failover protection.

Command

Consistent trustworthy answers.

Use current hardware and extend existing skills sets.

Nurture innovation for production.

Figure 1: How personas interact with data and analytics

(WebDAV). A full list of key features is available in [1] and describes some of the capabilities that, when present, help reduce challenges for user collaboration. Figure 1 is a diagram that tells the story of ways in which information acts as inputs and outputs that move through various personas. When integration technology is absent, collaboration between these personas runs into roadblocks.

An additional challenge is the reliance upon System Integrators (SI) whose “black box” solutions limit users outside of the SI team’s involvement, thereby inhibiting collaboration. These integrators also tend to become competitors because they can draw on open source tools and engineer solutions that entirely cut-out commercial capabilities.

A final challenge that is emerging is a paradigm shift in how some analytic tools are used in practice. As more routine and manual analytic work is being automated, the user is becoming more involved in improving the technology that performs the automated processes. An example is involvement in training machine learning models. This means the user activities are evolving, where some burdens are alleviated but new ones are added and require instruction.

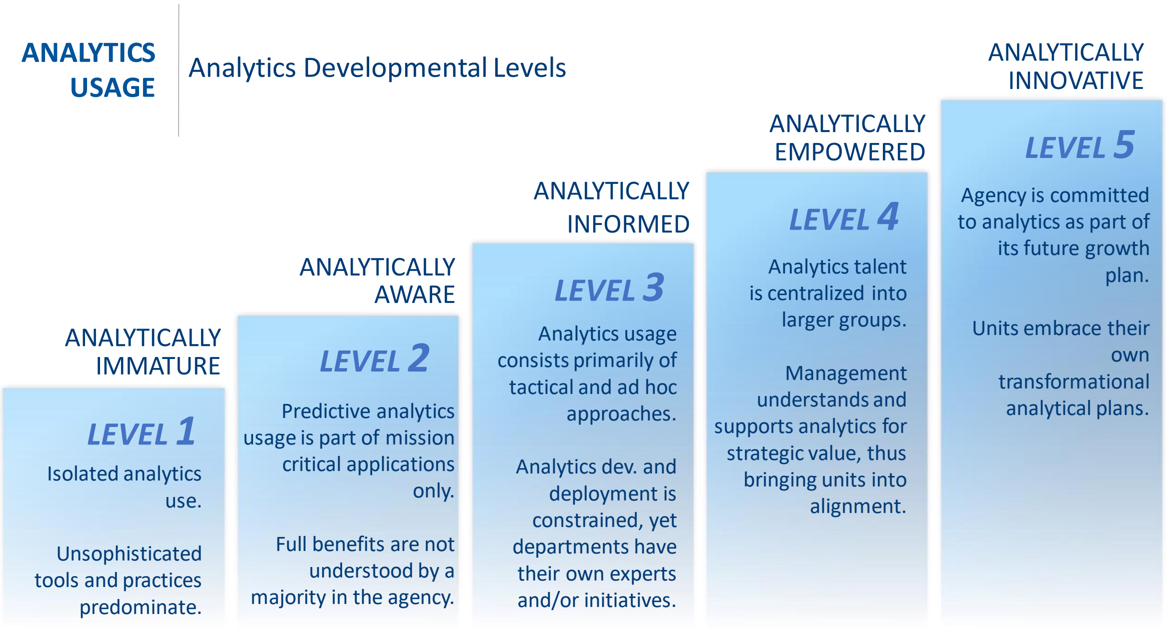

ANALYTICS USAGE
Analytics Developmental Levels
ANALYTICALLY IMMATURE
LEVEL 1
Isolated analytics use.
Unsophisticated tools and practices predominate.
ANALYTICALLY AWARE
LEVEL 2
Predictive analytics usage is part of mission critical applications only.
Full benefits are not understood by a majority in the agency.
ANALYTICALLY INFORMED
LEVEL 3
Analytics usage consists primarily of tactical and ad hoc approaches.
Analytics dev. and deployment is constrained, yet departments have their own experts and/or initiatives.
ANALYTICALLY EMPOWERED
LEVEL 4
Analytics talent is centralized into larger groups.
Management understands and supports analytics for strategic value, thus bringing units into alignment.
ANALYTICALLY INNOVATIVE
LEVEL 5
Agency is committed to analytics as part of its future growth plan.
Units embrace their own transformational analytical plans.

Figure 2: Analytic Usage Maturity Model

Concept and Initial Studies

Given the identified motivations of government and industry, the Roundtable wrote several studies to help define a concept to meet the goals. These studies included an examination of analysis tools, techniques and procedures (TTPs) along with use cases where they could be applied [2], an assessment of vendor perspectives on trends in government acquisition of analytic tools and data management architecture [3], a survey of specific industry analytic capability offerings [4], and the landscape of available and preferred analytic architectures across government agencies [5].

Analysis Tools, Techniques, and Procedures

The Roundtable's approach to examining TTPs applied the concept of a maturity model with developmental levels is shown in Figure 2 (adapted from Thomas Davenport and Jeanne Harris' *Competing on Analytics* [6] and Evan Stubbs' *The Value of Business Analytics* [7]). This can help articulate patterns of how they are combined to achieve objectives in different contexts. An organization can only apply TTPs where it has the skills, capabilities, and capacity to deploy, and this chart provides a perspective for understanding an organization’s current asset allocation and priorities and what barriers exist to achieving higher levels of analytic adoption. Otherwise, the organization shifts their effort from its mission to leveraging these TTPs, often without any success. Instead, recognizing where an organization exists in the maturity model can help it empower its analysts instead of overwhelming them [2].

While an organization might have traits in different developmental levels, the levels proved helpful in recognizing the requirements across use cases. The study that examined the TTPs from this perspective also articulated several example environments where the consequences of different degrees of analytic immersion were weighed to express the distinctions between the development levels and their requirements, costs, and scope. The key observation was that discerning the insights needed for mission focused analytic activities can apply across use cases whether in a government or commercial context because the essential elements of an analytic or investigative solution are effectively the same across those contexts [2].

Industry Architecture Survey and Review

The Roundtable performed an assessment in FY16 of product vendor perspectives on the trends they had witnessed in government acquisitions that specifically related to data management architectures and analytic tools. A key finding was that the government clients demonstrated an increasing tendency to choose commercial technologies over developing their own software, preferring those solutions that can be customized instead of performing their own custom software development. Cost is emerging as a potent influence with decreasing budgets, but the openness and extensibility of the solutions is also important to the government [3].

Examining this from the vendor perspective as well, the survey found that vendors identified time-tested, matured platforms as being successful and preferred over the “bleeding edge” capabilities that push the limits of processing and storage technology, sacrificing reliability and robustness. Furthermore, vendors also emphasized the importance of openness and extensibility in their solutions [3].

Industry Analytic Capability

The Roundtable performed a survey in FY16 of the scope of analytic tools built and offered by commercial software vendors, defining different "archetypal" categories of analytics to help organize the varied tools, even if there is some overlap in capability. These classes included text analytics, structured descriptive analytics, predictive analytics, geospatial analytics, link and network analytics, streaming analytics, imagery analysis tools, audio analysis tools, and biometrics analytics. The study also defined each, listed specific capabilities that fell into these categories, and provided representative examples. Ultimately, this work also characterized the importance of these different capability categories to a variety of use cases (e.g., cyber defense, counter-narcotics), indicating which were critical capabilities, which were of different levels of moderate importance, and which were optional due to less relevance or a lack of data [4].

Landscape of Government Analytic Architectures

As a final study, the Roundtable also explored the state of analytic architectures within the US Government across different agencies. Given the diversity of needs and missions across government agencies, many existing solutions were encountered that handled different scales of data, but no agency had a settled solution and no overarching solution spanned all agencies. As part of this study, the Roundtable made an initial proposal for a "common exchange service," which would could be used by government and industry as a hub for collaboration while being relevant to different use cases and accommodating to different analytic technology implementations. This core idea eventually became what the Roundtable pursued in its pilots: the Analysis Exchange Model, described in greater detail below [5].

Technology

To address the challenges and achieve the benefits that underlie government and industry’s motivations, the Roundtable engineered the Analysis Exchange Model. This was created to provide a reference vocabulary for consistent reusable exchanges of information and analysis products, while upholding certain principles:

- Remain tool agnostic
- Be executable anywhere
- Be portable to any Information Technology environment

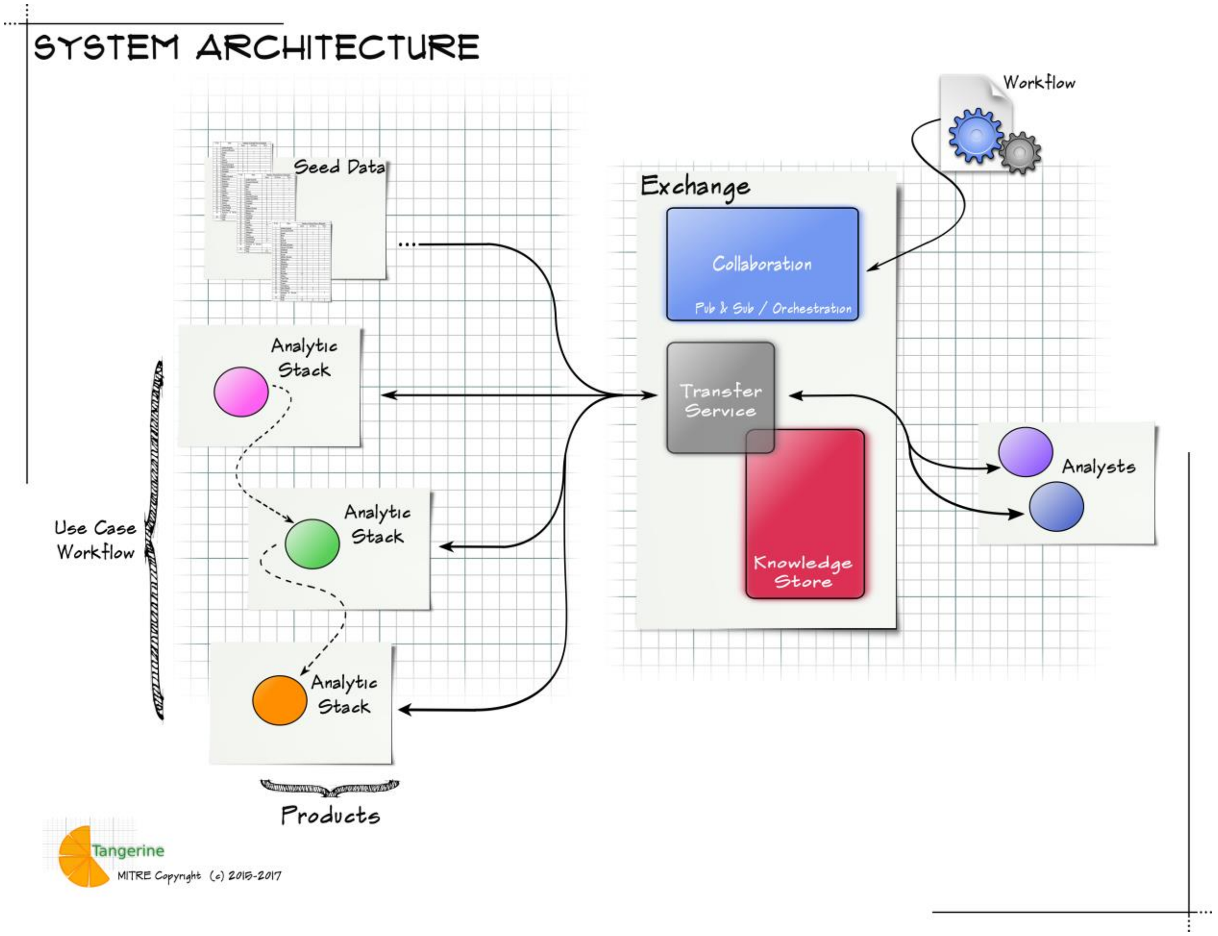


Figure 3: Exchange for automated analytic collaboration across industry and government.

- Be capable of management and inventory of all analytic assets

Figure 3 depicts the layout of the key architectural components invented for the Analysis Exchange Model 1.0. Its most important design elements include the Analysis Exchange Ontology (the defining representation of what is shared), the Knowledge Store (the inference and query engine and its backing store), and the Transfer Service (externally facing APIs and adapters). Any number of analytic stacks, possibly representing a variety of industry analytic capability, can interact in a workflow operating on data and producing knowledge when following this architecture.

Analysis Exchange Ontology

Paramount to success in using the Analysis Exchange is allowing for integration across the results provided by the varied analytic tools interacting in a workflow. These results tend to follow their own semantic models or at least have precise definitions, but they rarely achieve an exact match with one another in terminology and meaning. Therefore, the Analysis Exchange Ontology was developed to be a common, publicly known representation to which these results could be adapted in a mapping. While defining a complete ontology for all possible use cases was outside the scope of our work, we designed the Analysis Exchange Ontology to capture fully the elements related to the use cases we explored while also being extensible to any new use cases' required entities, relationships, properties, and rules. It was essential for our ontology to handle varied and potentially unrelated domains, addressing any ambiguities or incompatibilities that arose when collisions in terms were encountered.

The Analysis Exchange Ontology divided its hierarchical levels into upper, middle, and lower. Upper elements are those universal across all knowledge domains and provide the superstructure of the ontology, while the middle layer includes more concrete entities expected to appear across multiple domains, and the lower ontology consisted of entities typically specific to a use case. A well-known example of an upper ontology is Basic Formal Ontology (BFO) [8], which is small relative to many other ontologies but intended to be extensible into multiple domains. Many of the middle level entity types that appear in the Suggested Upper Merged Ontology (SUMO) [9] span the Analysis Exchange Ontology from the upper level down into the middle level. Despite the complexity of the ontology, it is common that most use cases are concerned only with middle and lower level elements and usually only a small subset of the overall ontology.

The Analysis Exchange Ontology is maintained in Web Ontology Language (OWL) that defines the entities, relationships, properties, type hierarchy, and inference rules. Statements and rules conforming to this ontology used by the Analysis Exchange are encoded in a Flora-2 based F-Logic. Adapters consume ingested content and analytic results to populate the Analysis Exchange's Knowledge Store with instances of knowledge conforming this model.

Knowledge Store

The Knowledge Store is where analytic results submitted to the Analysis Exchange are retained, either permanently or temporarily depending on the needs of a use case. In the Analysis Exchange Model 1.0, MongoDB was the basis for the store and the knowledge is represented in an F-Logic format that aligns to the semantics of the Analysis Exchange Ontology. The organizational convention is a collection per ingested record, where a record is an instance of contextually connected entities independent of the other records. A consistent naming convention across the collections aids in tracking the creation and content of each.

Transfer Service and Adapters

The Transfer Service handles transport and adaptation activities. As far as transport, the Analysis Exchange application programming interfaces (APIs) are how analytic stacks and end users access the Analysis Exchange, allowing the upload and retrieval of analytic results. For Analysis Exchange Model 1.0, these RESTful APIs, or web service APIs, are written in Java and employ a standing Apache Tomcat server.

Adapters address the challenge of the disconnect in formats and models of analytic results across different industry analytic tools. To interoperate, these analytics require a common language, which is the purpose of the Analysis Exchange Ontology. Derived from a concept articulated in the MOSAIC project [10, 11], which dealt with combining artifacts of loosely coupled analytics that were not originally engineered to communicate, adapters convert results into a format conforming to the Analysis Exchange Ontology and vice versa. This also means that there must be adapters for each analytic result type used.

The principal work required in engineering an adapter for a specific analytic format is understanding the semantic alignment between it and the Analysis Exchange Ontology, namely how content in the analytic format should map to what is specified in the Analysis Exchange Ontology. Mistakes in the alignment can lead to errors propagating to the Knowledge Store and through a workflow, so a cautious approach to this work is required, making this one of the more intensive efforts as new analytic formats must be included in a use case. In most cases, someone only interested in a specific use case can constrain what they are interested in preserving from their analytic results to a subset of what is covered by the entire ontology and can limit the mapping of the related adapter to only the relevant elements of the Analysis Exchange Ontology.

In the Analysis Exchange Model 1.0, adapters are written in Java and data mappings that direct adapter behavior are composed in JavaScript Object Notation (JSON).

Use Cases

For two years, the Analysis Exchange Model was explored for a variety of impactful use cases that could demonstrate 1) how it could draw on the collective analytic power of multiple industry partners' solutions and 2) the potential impact of breaking down barriers to collaboration. This section provides an overview of these different use cases.

FY17 – Fraud, Waste, and Abuse

The federal government distributes reimbursements for travel to thousands of recipients each year, making it a target for fraud, waste, and abuse. Data analysis drawing on different tools can help identify and defeat those attempting to exploit the government's good faith. In this use case, the Analysis Exchange Model served as a consistent structure for information drawn from multiple sources relevant to travel vouchers, where different analytic tools would retrieve the data from the Analysis Exchange, generate their enrichment, and store this as an addition to the records. The contents of free text fields in the travel voucher undergo content extraction to identify parts of names and addresses. Location content discovered in these fields are geolocated with a different analytic that also determined routes between the stated destinations in the travel voucher. Personally identifiable information (PII) data drawn from providers with access to this add more features to a record. Then statistical analysis is performed to find outliers in the data, which is visualized in a graphical representation. Each of these capabilities was provided by a different Roundtable partner without the need for direct connectivity from beginning to end. This serves as evidence that the Analysis Exchange therefore empowers analysts to use all the tools at their disposal to identify the perpetrators of fraud [12].

FY17 – Threat Assessment

Through proactively identifying, assessing, and mitigating threats, threat assessments can lower the risk of targeted violence, but they are costly to perform both in time and money. Analytic software capabilities reduce these costs and can enable analysts to discover patterns that indicate current and emerging violent events. In this use case, a variety of tools were brought to bear against the Armed Conflict Location and

Event Data Project (ACLED)[2] corpus. The partners enhanced this openly available data by drawing on their own collected and curated data, with natural language processing (NLP) used to provide sentiment analysis against the relevant unstructured textual content. This is ultimately visualized with drill charts, timelines, link analysis, and geospatial presentation from multiple partners [13].

FY17 – Cyber Forensics

Hackers pose cyber threats that continuously defeat security systems and challenge cyber security experts. A forensic study of who perpetrated an attack and the extent of the damage done is a high priority for analysts, but given the rapidity of how cyberattacks occur, the same analysts also want to proactively monitor networks for threats to discover and stop them. Either scenario requires drawing on a superfluity of data from different sources, which led to the Analysis Exchange design guiding a solution that sifted and pooled information from these sources for this cyber analysis. In the case of the reactive (or forensic) analysis, the implementation explored answering the topics of the attribution of a cyberattack, the extent of its impact, the scope of data loss or breach, and remediation steps to follow the event. In the case of the proactive analysis, the emphasis shifted to a task of hunting threats where both threat and malicious actor are unknown, utilizing open source data to identify and judge threats and malicious actors who may be targeting a given organization [14].

FY18 – Army TRADOC Support

In the subsequent fiscal year, the Roundtable took on two use cases under the direction of government sponsors, US Army Training and Doctrine Command (TRADOC) and US Air Force A2.

For Army TRADOC, the problem space was to address the current process for generating reports and building out networks, a manual activity of drawing input from multiple sources to be processed, consolidated, and correlated. The effort's goal was to facilitate this activity with automation, making use of the Analysis Exchange design established by MITRE in the previous year to bring tools together to solve

[2] www.acleddata.com

this new problem. Army TRADOC also provided an Operational Environment (OE) ontology, which the Roundtable aligned to its original Analysis Exchange Ontology to augment its representational power and ensure critical elements to the use case were present.

Two separate teams approached this task. The goal of the first team was to automate the part of the workflow that dealt with building out networks of relationships, leveraging the textual intelligence reports that were the basis of creating these networks. Initial data undergoes entity extraction from the content using multiple tools from different vendors, adapting these to the Analysis Exchange Ontology's knowledge semantics, and then ultimately visualizing the graphical relationships encoded in the original source, making this available to the analysts to examine. Similarly, the second team also performed content extraction, followed by social network analysis and geotagging to produce knowledge adapted into the Analysis Exchange Ontology.

In the following year, the Analysis Exchange work has evolved into an Army pilot described in the next section.

FY18 – Air Force A2 Support

Before engaging the Roundtable, Air Force had sponsored an innovation initiative to develop an information architecture for integrating analysis, collection, and targeting activities across the intelligence, surveillance, and reconnaissance (ISR) enterprise. The focus was on creating a USAF object-based production (OBP) model for data indexing and correlation with other US Intelligence Community (IC) and US Department of Defense (DoD) structured object architectures. Working with the Roundtable, the Air Force requested a pilot drawn from industry solutions to build and manage an ontology, create analytics to discover and extract ontology-relevant data from unstructured content, structure content into the ontology, perform analysis of this structured content, and query dynamically against this content. The Air Force emphasized a desire to have a method where partners collaborate in a shared environment that supported the ability to prototype and refresh analytic tools.

Four teams approached this topic. The first two teams took an approach that aligned with the approach to the Army TRADOC use case but drawing on different data sources offered by Roundtable partners that were data providers. These teams took an end to end approach to the work, but two other teams examined the use case from 1) the perspective of the advance work of content extraction and analysis before ingest into the shared environment, drawing on multiple tools and aligning their results to the Analysis Exchange, and 2) the perspective of downstream visualization and representation of content previously structured and shared in the Analysis Exchange.

FY19 Pilot for Army TRADOC G2

In FY19, the Roundtable has taken on an effort to realize the Analysis Exchange capability for United States Army TRADOC G2. This pilot seeks to deliver a product that will ultimately be a baseline for workflow and an opportunity for deploying analytic capability via the cloud. The intent is to reduce the time, resources, and overall burden of data preparation and triage and, at the opposite end of a given workflow, help analysts more effectively and quickly act on emerging situations. The Roundtable partners provide the cloud infrastructure, the analytic capability, and the visualization, weaving these together following the pattern established in the previous years' Analysis Exchange efforts.

Two desired workflows were selected by Army TRADOC and the Roundtable for the pilot. The first of these workflows, Global Events Watch, requires the identification of specific, imminent events in the OE, recognizing where they are expected to occur, and analyzing them for their likely effects. The other workflow pursued, OE Trends Forecasting, takes a longer view than Global Events Watch. This workflow requires the production of historical trend lines with impacts that will be felt in the near-term (5 years out) and mid-term (10 years out) future OE.

As input to the pilot use cases, the Roundtable's industry partners draw on open source data, both structured (such as ACLED), unstructured (open source regional news, scientific literature), and mixed (GDELT).[3] Army has also provided publicly released G2 reports and lists of external websites analysts frequently leverage to serve as additional input. As output, the pilot will offer a dashboard for analysts to examine the content, providing summaries of the text, geographic plots of events, and notable and identifiable changes in the OE. In addition, for the OE Trends Forecasting, trend lines of scores and confidences of these trends are also examined, with an initial emphasis on topics related to urbanization and demographics and climate change. Ultimately, Army's goal is to have a workflow that can serve as a real framework to allow the flow of data between the tools to produce the desired results.

[3] https://www.gdeltproject.org/

Metrics

The benefits incurred by using the Analysis Exchange as a solution require metrics to evaluate its success, and we consider Army TRADOC's needs and metrics as a case study. Certain questions about metrics were raised and Army TRADOC authored answers and their insights into evaluation.

Question: What analytic processes that were once manual would now be automated by an analytic software tool? What savings does that allow in terms of time, money, and resources?

Response: Automating the data generation process for training scenarios would reduce the manpower and time it would take to develop the data package. This reduction results in an opportunity cost increase to support additional training units.

Question: What tangible metrics are there for judging whether the Analysis Exchange has an impact?

Response:

- *Time-to-market for our data products*: we anticipate reducing the time from 2 weeks to 2 days.
- *Units served*: we anticipate having an increased capacity to support twice as many units in a given year.

Question: Does this represent the ability to do current tasks more efficiently vs. the ability to do new tasks previously infeasible?

Response: This represents the intelligent automation of existing processes. We also see significant opportunity to explore tasks and new processes to increase the level of both depth and complexity in our analytic product.

Impact

The Technology & Innovation Roundtable has striven to achieve certain goals in their ongoing work on the Analysis Exchange Model. At a high level, these goals represent mutual benefits to both government and industry. First, the Analysis Exchange supports a culture of consistent and repeatable interoperability between analytic tools, which allows them to compete more efficiently. Second, the Analysis Exchange allows for meeting the government needs with a lower investment of resources. Third, the Analysis Exchange represents a model that can supplant the data silos that emerge when multiple analytic tools are developed or used in isolation. Fourth, the Analysis Exchange provides the basis for a consistent standard for the analytic community which can be freely reused, allowing them to avoid the expenditures associated with inventing this anew each time. Finally, the ability to draw on multiple analytic solutions with the Analysis Exchange leads to creating enriched products based on the best market solutions that can be continually refreshed as new solutions are engineered, ensuring the pace of excellent overall results keeps up with the standards of all tools available. The interests of government and industry often align, and the Roundtable has provided a solution that serves both parties.

References

[1] SAS, "SAS Integration Technologies," 2011. [Online]. Available: https://www.sas.com/content/dam/SAS/en_us/doc/factsheet/sas-integration-technologies-101497.pdf. [Accessed 2019].

[2] A. Etches, C. Brown and J. Stultz, "Analytics and Use Cases," 8 November 2016. [Online]. Available: http://www2.mitre.org/public/analytic-technology/pdfs/Analytics_and_Use_Cases_Study_IBM_SAS_11_25_16.pdf.

[3] B. Choung and M. Chandler, "Review and Study on Industry Collaboration," 8 November 2016. [Online]. Available: http://www2.mitre.org/public/analytic-technology/pdfs/Industry_Architecture_Survey_and_Review_Study_Palantir_11_25_16.pdf.

[4] W. Niehaus, B. Adams, S. Panzer and L. Hanson, "Alignment of Capabilities," 8 November 2016. [Online]. Available: http://www2.mitre.org/public/analytic-technology/papers.html.

[5] R. Winder, J. Jubinski and A. O'Hanlon, "Analytic Architecture Survey and Review," 8 November 2016. [Online]. Available: http://www2.mitre.org/public/analytic-technology/pdfs/Architecture-Survey-Review.pdf.

[6] T. H. Davenport and J. G. Harris, Competing on Analytics: The New Science of Winning, Harvard Business Review Press, 2007.

[7] E. Stubbs, The Value of Business Analytics: Identifying the Path to Profitability, Wiley, 2011.

[8] R. Arp, B. Smith and A. Spear, Building Ontologies with Basic Formal Ontology, Cambridge, MA: The MIT Press, 2015.

[9] I. Niles and A. Pease, "Towards a standard upper ontology," in *Procedings of the International Conference on Formal Ontology in Information Systems*, 2001.

[10] R. Winder, J. Jubinski, J. Prange and N. Giles, "MOSAIC: a cohesive method for orchestrating discrete analytics in a distributed model," in *International Conference on Application of Natural Language to Information Systems*, Manchester, UK, 2013.

[11] R. Winder, N. Giles and J. Jubinski, "Implementation Recommendations for MOSAIC: A Workflow Architecture for Analytic Enrichment," McLean VA, 2011.

[12] C. Brown and J. Stultz, "Fraud, Waste, and Abuse Use Case," Analytic Technology Industry Roundtable, 2017.

[13] B. P. Adams, F. McGuinness, N. Riccio, L. Hanson, W. Niehaus, C. Walker and B. Biesecker, "Analytic Technology Industry Roundtable Threat Assessment Use Case," Analytic Technology Industry Roundtable, 2017.

[14] A. Etches and C. W. Walker, "Cyber Use Case," Analytic Technology Industry Roundtable, 2017.