

# **Analysis Exchange Framework**

## **Terms of Reference**

**December 2016**

Approved for Public Release; Distribution Unlimited. Case Number 16-4653

The views, opinions and/or findings contained in this report are those of The MITRE Corporation and should not be construed as an official government position, policy, or decision, unless designated by other documentation.

©2016 The MITRE Corporation. All rights reserved.

This page intentionally left blank.

## Background

The Analytic Technology Industry Roundtable has produced several studies focusing on analytic architectures, in order to gain a better understanding and ground truth for collaboration opportunity. The Roundtable produced studies on industry analytic capabilities, and use cases in 2016 [1, 2, 3, 4]. These studies can benefit the government while encouraging collaboration across industry partners. In this paper, we articulate the core architectural components of an Analysis Exchange Service<sup>1</sup> intended to support the desired collaboration. The exchange design is based upon several months of engagement, study, and work with the Industry Roundtable Working Group. This paper describes the common exchange that the Roundtable will build. The paper describes how MITRE and industry will engage to realize this model. In order to achieve a successful prototype for the Analysis Exchange Service and a supporting Analysis Exchange Model, industry and MITRE will be working together in partnership with government input.

- *Analysis Exchange Service* – a network-enabled repository for results from external analytic services, fostering industry and government collaboration by facilitating knowledge fusion and downstream analysis.
- *Analysis Exchange Model* – the ontology capable of representing adapted analytic results from various sources and intended to support the use case(s) generated for the prototype Analysis Exchange Service.

This document expands on these components and describes the path toward implementing them. This will be a key Roundtable effort into FY17. This document also articulates the details about what Roundtable’s tasks will be (Figure 1) and why. Government also will have a role in this effort, specifically in providing strategic input and long-term requirements for how it can evolve to solve their hard problems.

### Tasks

- Use case maturation
- Analytic tool selection
- Data selection
- Exchange service APIs
- Publish-subscribe service
- Data model / ontology
- Semi-persistent result store
- Adapters
- Analytic output guidance
- Testing the workflow

*Figure 1*

---

<sup>1</sup> In our original material, this was referred to as the *Common Exchange Service*.

## Charter

This study will:

- Provide an architecture to support collaboration, whether through a publish-subscribe service or a more involved orchestration of externally based analytics.
- Provide a data model with the supporting ontology that can be publicly released and serve as a lingua franca for analytic results (this data model, the Analysis Exchange Model, is described in more detail in the next section).
- Provide the templates, or at least the pattern of activity, to allow for the adaptation of different industry analytic results into a common data model.

## Study Products

Each of the goals specified above introduces different products that require development, either from scratch or by applying and extending existing capabilities. These include the following:

- The application programming interfaces (APIs) to allow for interaction with the Analysis Exchange Service.
- A publish-subscribe service that acts as a basis for the exchange and can be the foundation for orchestration.
- An ontology serving as the underlying data model and rules for reasoning for the Analysis Exchange Model.
- A semi-persistent store for the analytic results mapped into the Analysis Exchange Model.
- Template adapters for migrating analytic results into the Analysis Exchange Model.

There are also several dependencies of note, including:

- Maturing and finalizing the use case.
- Selecting analytic tools, systems, and capabilities for the use case, which represents the industry collaboration at the tool level.
- Selecting and collecting the input data sets for the use case (for example, social media data, images, or plaintext).
- Assisting and guiding the creation of the adapters between analytic results and the Analysis Exchange Model.
- Testing, evaluating, and studying the analytic workflow in how well it fulfils the use case.

The remainder of this document explores the charter goals and planned study products in greater depth.

## Analysis Exchange Service

The Analysis Exchange Service (as depicted in Figure 2) is a network-enabled repository and acts as a hub between the automated analytic stacks from different data and analytic providers and the analyst customers who make use of the artifacts, knowledge, and results of the analytic processes. The goal of the Analysis Exchange Service is to achieve fusion of the artifacts and knowledge it holds so that different industry automated analytics can contribute to a final product without compromising their internal analytic technology. Further analysis of what the Analysis Exchange Service holds is possible downstream of this service. In all, the Analysis Exchange Service will encourage new collaborative capabilities between industry partners in the service of their government customers.

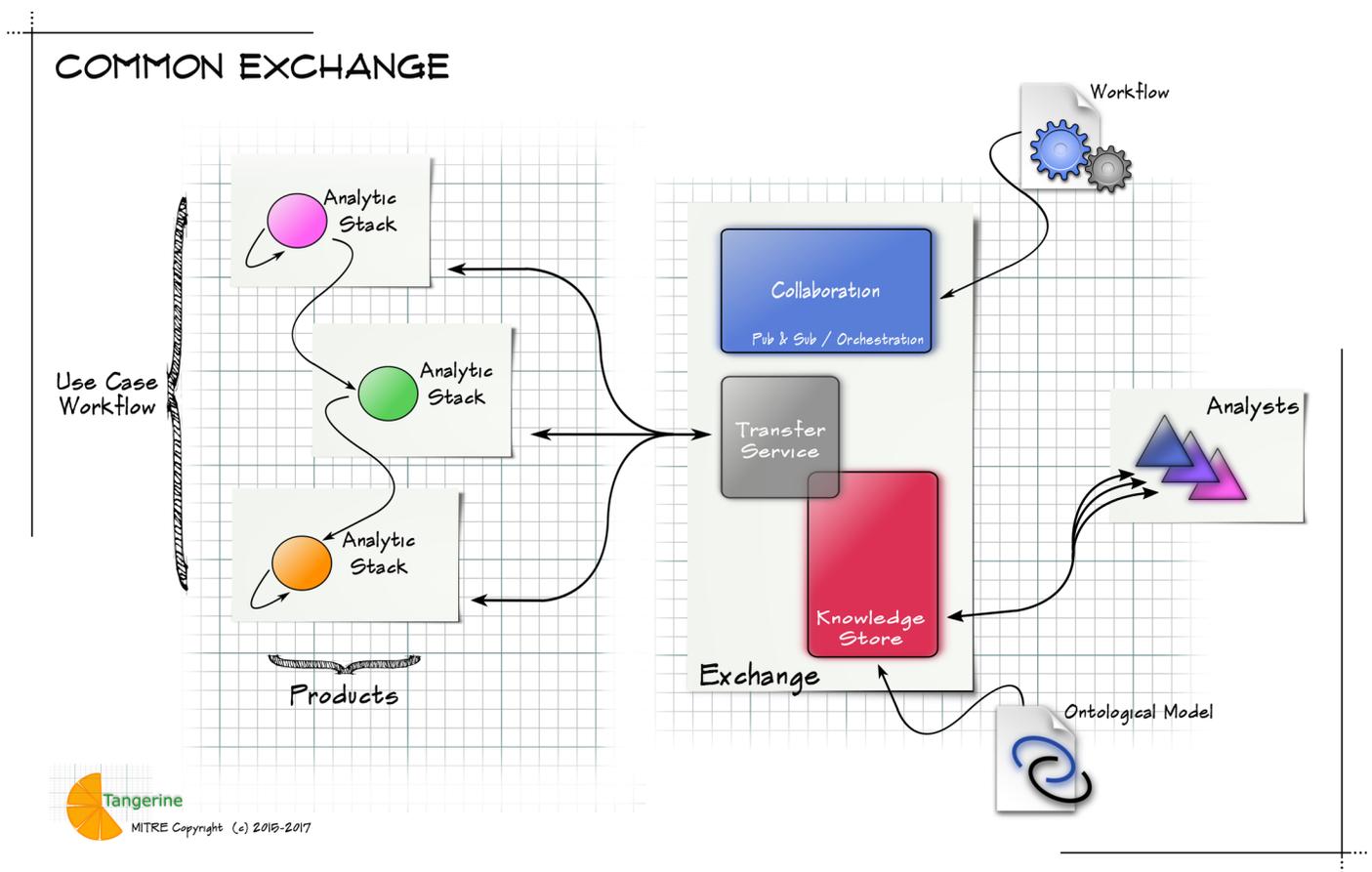


Figure 2: Exchange for automated analytic collaboration across industry and government.

As specified above, there are necessary components to develop as part of the Analysis Exchange Service. The application programming interfaces, which correspond to the transfer service depicted in Figure 2, are the entry point into the architecture that can be controlled following an ad hoc publish-subscribe pattern; these can also be extended to allow workflow orchestration to call on external analytics—and potentially internal analytics eventually. The internal store retains artifacts for as long as the use case requires. Further, the results in this store are defined by a consistent data model that conforms to an ontology built for the supported use case, which seeks to mitigate ambiguities between different representations that the external analytics might exhibit. Adapters are required to make the format and possible semantic conversion into this model, which ensures that downstream analysts receive a consistent picture of the results regardless of the source.

Activities that are supported by the publish-subscribe model include 1) *pipelines* of analytic activity that execute when informed through the subscription that the results they expect are available in the repository, 2) *aggregation* from multiple sources into a final product, and 3) *alteration*, or updates, of results previously submitted. These activities align to the data flows depicted in Figure 3, as first expressed in the MITRE BlueRidge project.

Later sections will expand on how these pieces fit together and how each will be implemented over the course of FY17.

## Analysis Exchange Model

The Analysis Exchange Model is the ontology that defines the space of analytic results which are adapted to conform to it. This aligns to elements that recur throughout enterprises where analytic architectures are being explored, namely the need for a common representation and consistent definition of knowledge objects that make up the results. While the Analysis Exchange Service can store results in their raw output form, an Analysis Exchange Model and an evolving library of adaptation software that keeps pace with changing native models means useful interactions can spin up more quickly. Additionally, it does not assume the burden of adaptation necessarily has 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

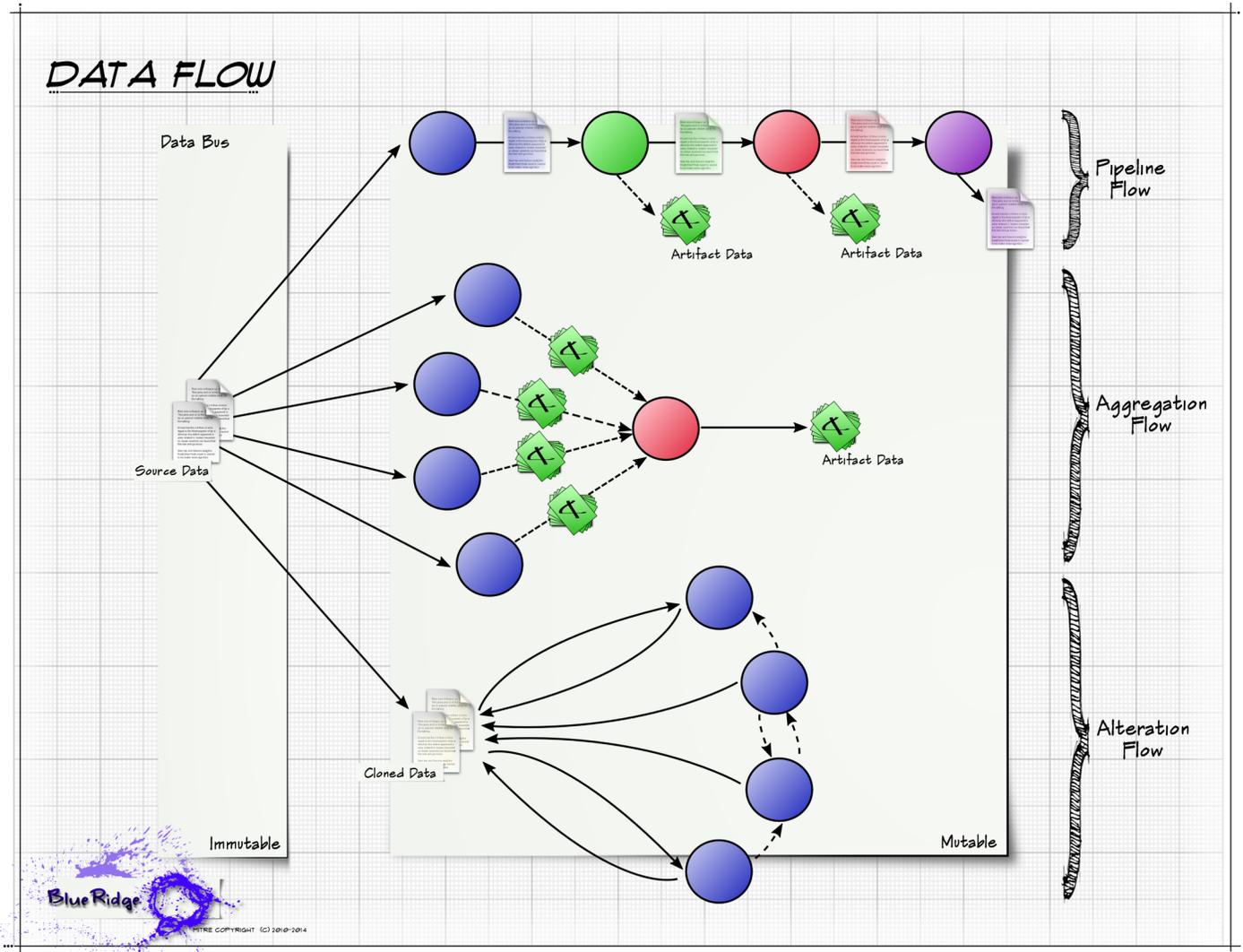


Figure 3: Flows of data (or results) as originally presented in MITRE BlueRidge work

Model. In this case, we intend to design the Analysis Exchange Model and representative adapters to establish the template for future adapter development.

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 preferable to 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 might serve as an upper ontology, but we plan to create a tailored, more domain-specific ontology.

## Components

### APIs for the Analysis Exchange Service

The principal goal of the Analysis Exchange Service is to bridge capabilities between different industry partners in service of government customers, in particular analysts. The application programming interfaces for the Analysis Exchange Service are the principal ways industry partners and government customers will be able to access the Analysis Exchange Service, covering activities such as uploading results, collecting results to ingest in downstream automated analytics, and retrieving the final results that represent the outcome of a complete activity, whether as an instance of a repeatable workflow or a single ad hoc experiment bridging multiple analytic capabilities.

It is part of the tasking to create all the necessary APIs to perform the actions described above. These APIs can be the foundation of providing a publish-subscribe service or more robust centralized orchestration that calls on external analytics to execute. These APIs will be written in Java, but they can be extended to other languages based on a language-independent RESTful implementation. Eventually, there can also be a capability in the APIs to allow for partners and customers to execute internal analytic processes in the Analysis Exchange Service, but the initial vision expects the analytics to be those provided by industry partners and to remain external.

### Publish-Subscribe Service and Orchestration

Following a publish-subscribe pattern for communication is a straightforward method that imposes a minimal burden. This model expects that when analytic artifacts are made available, the Analysis Exchange Service will publish that these artifacts. Therefore, subscribers awaiting certain expected analytic artifacts will be notified when the artifacts that interest them, either for a single experiment or part of a regular workflow, are available.

It is part of the tasking to implement this publish-subscribe service. While a publish-subscribe pattern typically assumes an underlying structure of clients and a central server, we can use a framework<sup>2</sup> to change this structure into one that is peer-to-peer. This allows it to act as a messaging service between the Analysis Exchange Service and the external systems when updates occur. We expect that partners will subscribe to the artifacts published by the Analysis Exchange Service as well as provide the artifacts that the Analysis Exchange Service makes available. Who acts in what capacity should be agreed upon as part of defining an initial use case. Also part of this task is developing a listener that ensures events are noticed, where events are any kind of relevant notification about a change in the system or data or any requests that are part of a workflow.

Eventually, this publish-subscribe service can also provide the basis for more complex internal orchestration, where a centralized control calls on both external and internal analytics to execute on different stages of artifacts to produce the final results. The initial vision intends to keep this simple, but with the potential to be extended to greater responsibilities held by the Analysis Exchange Service, perhaps using an open source scientific workflow application as the basis for the orchestration. The listener we described in the publish-subscribe pattern could also be used in this more central orchestration.

## Ontology and Data Model

An underlying need in a successful Analysis Exchange Service is achieving a seamless integration across varied analytic results in support of a use case. These results are often in semantic models or can be made to conform to them. We envision applying an ontology (namely a set of defined classes, relationships, and rules) that describe the elements relevant to the use case, which allows for a consistent treatment of the knowledge and the capacity for reasoning across the rules that describe how different statements of knowledge affect one another.

Developing this ontology, or the Analysis Exchange Model as described above, is one of the main tasks of Roundtable effort. This ontology will be integrated with the store that will be populated with instances of knowledge by the adapters.

---

<sup>2</sup> For instance, Apache Kafka

The store and adapters are both described below. The principal dependency here is a clear use case. While we can begin design and make changes to fit an evolving use case, having a direction of the space the resulting knowledge occupies is important to know what to include in the ontology's classes, relationships, and rules and what to exclude as irrelevant.

## Store

While analytic artifacts published by the Analysis Exchange Service conform to the Analysis Exchange Model, they also need a place to be stored internally. This store can be permanent, volatile, or somewhere in between depending on the use case. In some cases, all that is of interest is timely data whereas in others, aggregation over successive uploads provides the needed picture to downstream analytics and analysts. There can also be degrees of volatility, where all knowledge is at least retained temporarily and some is always retained long term. Regardless, this suggests a knowledge base that can potentially scale if considerable artifacts are produced.

Implementing this knowledge store is a task to be performed in concert with the ontology development. We expect this requires both a reasoning engine and persistent store. The length of retention will be determined by what fits the use case.

## Template Adapters

One of the more serious challenges to inter-industry and government collaboration is the disconnect in formats and models of analytic results. In order for analytics to operate in concert, whether they ingest one another's results as input or synthesize a final product collectively, they must have a common language. This is what the Analysis Exchange Model does, but results must be mapped into (and in some cases out of) the Analysis Exchange Model, as we know that we cannot mandate that every analytic will produce (or consume) our model. This is where adapters become essential. Adapters convert the results from their raw analytic results into a format conforming to the Analysis Exchange Model and vice versa if the use case dictates. This also means that there must be adapters for each analytic involved, or at least each analytic result type involved (formats can potentially span different analytics).

Developing and implementing template adapters is another task. While adapters can be bi-directional (going from raw analytic results to a format conforming to the Analysis Exchange Model and from the Analysis Exchange Model to a format ingestible by external analytics), we intend to only start with adapters that map from analytic results into the Analysis

Exchange Model, unless the use case requires otherwise. The intent is to make the adapters necessary to execute and demonstrate the use case; these adapters can eventually become the basis for the continual development of more adapters for other analytic results. This requires a use case and ongoing collaboration to ensure we understand the analytic output and are preserving the semantics in the adaptation.

## Conclusion

Achieving both the Analysis Exchange Service and the Analysis Exchange Model requires simultaneous development and implementation by the Roundtable of several different elements, including the APIs, the publish-subscribe service, the data model and its underlying ontology, a semi-persistent store for results, and adapters. Each of these ensures that the Analysis Exchange Service can be a central hub for collaboration, but it also levies certain dependencies that have been discussed throughout the proposal. These dependencies include maturing and finalizing the use case, selecting analytic tools, selecting the input data sets relevant to the use case, testing the workflow to ensure it fulfills the use case, and ongoing guidance of the development of the Analysis Exchange Service, particularly the adapters which must properly preserve the content of the analytic results.

Between the tasks and the dependencies, the goal of achieving an analytic architecture fostering collaboration can be achieved. This can also be expanded and extended in the future by following the process when applying it to new use cases or further enriching the control of the Analysis Exchange Service by allowing it to handle orchestration centrally.

## References

- [1] 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](http://www2.mitre.org/public/analytic-technology/pdfs/Analytics_and_Use_Cases_Study_IBM_SAS_11_25_16.pdf).
- [2] 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](http://www2.mitre.org/public/analytic-technology/pdfs/Industry_Architecture_Survey_and_Review_Study_Palantir_11_25_16.pdf).
- [3] 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>.
- [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>.