



# Improving DoD Data Strategy

6 Ways Rosette by Babel Street Can Help

# Introduction

In 2020, the Department of Defense articulated a significant new data strategy. The strategy entails “accelerating the Department’s transition to a data-centric organization that uses data at speed and scale for operational advantage and increased efficiency.”<sup>1</sup>

Once implemented, the DoD strategy will improve and hasten analytical insight. This need is clear. In a world where information volume doubles biennially,<sup>2</sup> the influx of massive new datasets — coupled with the DoD’s siloed data-storage structure — has left analysts spending up to 80 percent of their time searching for information, and only 20 percent of their time analyzing it.<sup>3</sup> Understanding this, a recent Department of Defense Intelligence Information System’s (DoDIIS) Worldwide Conference focused on specific challenges the DoD must meet to successfully implement its data strategy.

We at Babel Street strongly support the DoD’s efforts. We understand that modern warfare requires decision making at machine speed. This type of decision making can only be accomplished by providing intelligence analysts with tools to derive the faster, deeper insight that comes from improved data access. To gain this data advantage, the DoD must implement

secure technologies to triage, assimilate, and enrich information across source languages.

These technologies must empower intelligence analysts to rapidly extract essential elements of information from text and share information with authorized partners. Software tools should also help analysts interact with machine learning models — training, annotating, and refining them as needed to obtain maximum insight, then deploying these models at scale.

Using artificial intelligence to fuel cutting-edge data analytics is Babel Street’s business. We believe that the Rosette® text analytics and discovery platform can help the DoD meet the challenges expressed in its data analytics strategy. Rosette employs natural language processing, machine learning, and statistical modeling to analyze unstructured and semi-structured text across hundreds of languages, quickly revealing valuable information and actionable data.

The rest of this ebook will examine how Rosette meets the DoD’s data and analytics challenges as expressed at the DoDIIS Worldwide Conference.

## 1. ISSUE

# Containerization and data protection as a service

### Challenge

DoD needs data intelligence technology that it can easily deploy across a broad range of enterprise data systems, while concurrently ensuring the protection of the data stored in those systems.



### How Rosette can help

Rosette is easily deployed and aligns with existing DoD data security protocols and standards.

As a vendor-agnostic API, Rosette employs a modular architecture ready for containerized delivery via Docker and other containerization platforms.

We understand that data protection is a significant issue for the DoD. Rosette operates on premises across the Defense Intelligence Enterprise, analyzing data *in situ* without moving or storing it. Because Rosette leaves data as and where it's found, the software does not conflict with existing data protection protocols or standards.

## 2. ISSUE

# JWICS modernization

### Challenge

DoD needs to apply multilingual data triage and enrichment capabilities to data as that data is being created and collected in the Joint Worldwide Intelligence Communication System.

### How Rosette can help

Currently, the DoD uses a number of different systems to collect and store data in the JWICS. This leaves intelligence professionals unable to search the JWICS effectively. In fact, these analysts often must resort to binary search processes that do not consider “degrees of truth” — or the likelihood that a piece of data matches a search query beyond match/no match parameters.

Rosette helps optimize search functions by embedding multilingual natural language processing capabilities into the ETL pipeline.

Without Rosette capabilities at the ETL stage, tasks such as name translation, name transliteration, entity extraction, entity resolution, relationship extraction, topic analysis, and salient-term analysis are left to the individuals developing user-facing applications. There is no reasonable expectation that these software engineers can or will apply multilingual data triage and enrichment to every application. Applying these capabilities as part of ETL processes is the only way to provide the optimal searchability and discoverability that underpins improved intelligence.

### 3. ISSUE

# Implementing all-source object-based production

## Challenge

The DoD needs a holistic, rather than siloed, view of its data.



## How Rosette can help

Historically, DoD data has been stored in, and disseminated through, the agency that produced it. This leads to siloed, hard-to-access information. Object-based production processes will change that. Object-based production reaches through data silos to create knowledge objects (or informational “buckets” in which all information about a person, place, thing, or concept are stored) that all authorized analysts can access. In doing so, it dramatically improves insight enterprise-wide. By establishing relationships between objects and creating new types of groupings (all the IP addresses associated with a potential terrorist cell, for example) object-based production can even help forge new types of insight.

Rosette aids in the move to object-based production. It transforms unstructured textual data into structured knowledge objects ranging from single data points to entire knowledge webs.

## 4. ISSUE

# Zero-trust implementation with improved sharing

### Challenge

The DoD needs to both limit data access to authorized users and enable sharing among the largest possible pool of authorized users.

### How Rosette can help

As the DoD moves from agency-centric to enterprise-wide data availability, zero-trust processes must be refined to constantly audit who is accessing what data. These processes must work in tandem with an operational imperative to share information across the broadest possible group of authorized users.

Rosette can help the DoD accommodate these seemingly contradictory goals.

First, by statelessly extracting knowledge as structured metadata from unstructured input, Rosette enables role-

based access to key elements of information. In doing so, Rosette ensures that each of today's users is credentialed to access a given piece of data today.

Further, Rosette's data classification systems enable improved sharing. Rather than classifying information document by document, Rosette divides documents into their constituent data elements, enabling governance at a granular level.

Think of it this way. Capt. Smith either has Top Secret clearance or he does not. In typical user-access scenarios, if Capt. Smith has Top Secret clearance, he can access documents classified as Top Secret. If he only has Secret clearance, he may not access Top Secret documents.

But suppose only a small fraction of the information contained in a document is classified Top Secret; the rest is merely Secret and within the limits of Capt. Smith's authority to access.

Rosette classifies each data element within a document. This process, sometimes called the "atomization" of data, enables Capt. Smith to access the data within a document that he is authorized to see, while denying him access to unauthorized information within the same document.

## 5. ISSUE

# Sharing intelligence with Five Eyes partners

### Challenge

The DoD needs to improve intelligence sharing with its Five Eyes intelligence alliance partners in the United Kingdom, Australia, Canada, and New Zealand.



### How Rosette can help

The atomization of data discussed previously can assist in FVEY sharing. Currently, an entire document may be classified Top Secret NOFORN because of just a few data elements. Therefore, that document may not be shared with FVEY partners. But if much of the data contained in the document is classified Top Secret (without the NOFORN designation) that information should be made accessible to FVEY partners, to the extent that FVEY intelligence analysts have Top Secret clearance in their countries. Rosette data atomization enables this type of data sharing across the FVEY alliance.

## 6. ISSUE

# Embedding algorithms into the analysis tradecraft

### Challenge

Empower analysts to improve insight by annotating and customizing machine learning models



### How Rosette can help

Rosette tools enable analysts to incorporate model annotation, training, customization, and adaptation into their workflows — often bypassing the need for software engineers' involvement in these tasks.

The rationale behind these capabilities is simple: subject matter experts know more about their areas of study than anybody else. Why shouldn't they be able to encapsulate and annotate information? Why shouldn't they have input into their own AI models, so that AI recognizes and returns the information they need?

Rosette's ease of use makes it possible for subject matter experts to interact with AI models — training, annotating, enhancing, and refining them so that AI understands elements of text that the analyst recognizes as important, but that the model would otherwise ignore. These models learn from analyst input and provide immediate feedback on incremental improvement when finding the data the analyst needs. These capabilities improve analytical insight and analyst efficiency.



# Meeting the challenge

Fast, robust intelligence is crucial to United States defense. To improve intelligence, the DoD has undertaken a new data strategy. The Rosette text analytics and natural language processing platform meets the needs for data linguistics, categorization, name indexing, name translation, and identity/entity/relationship extraction discussed during the DoDIIS Worldwide Conference. In addition, we have an extensive record of successful engagements across the Defense Intelligence Enterprise, including with Federal law enforcement and the Department of Homeland Security.

We believe this combination of proven capabilities and experience uniquely qualifies us to assist the DoD in implementing its data strategy. Contact us to learn how Babel Street can help fulfill your mission.

## Endnotes

1. Department of Defense, "DoD Issues New Data Strategy," October 8, 2020, <https://www.defense.gov/News/Releases/Release/Article/2376629/dod-issues-new-data-strategy/>
2. Call for Code Daily, "The amount of data in the world doubles every two years," October 7, 2020, <https://medium.com/callforcode/the-amount-of-data-in-the-world-doubles-every-two-years-3c0be9263eb1>
3. Johnston, Catherine and Wright, Elmo C. et al, "Transforming Defense Analytics," Joint Force Quarterly, October 1, 2015, <https://ndupress.ndu.edu/Media/News/Article/621117/transforming-defense-analysis/>

## Babel Street. Unlock the most insights that matter.

Babel Street provides the most advanced data analytics and intelligence platform for the world's most trusted government and commercial brands. The AI-enabled platform helps them stay informed and improves decision-making for threat intelligence, risk mitigation, identity management, and alerting use cases. Teams are empowered to rapidly detect and collaborate on what matters in seconds by transforming massive amounts of global, multilingual data into actionable and contextual insights so they can act with confidence.

Learn more at [babelstreet.com](https://babelstreet.com)

