Analytic Technology
Industry Roundtable

Analytic Technology Industry Roundtable Threat Assessment Use Case

1. Introduction

1.1. Analytic Technology Industry Roundtable

The Analytic Technology Industry Roundtable brings together analysis and analytic technology companies to address industry challenges and discuss topics of mutual interest and concern. The Roundtable also engages with the U.S. government on selected topics to foster greater industry–government collaborations that can lead to better solutions.

1.2. Analytic Exchange

A priority project for the Roundtable is the creation of an Analysis Exchange (AE) that Roundtable companies can use to interconnect their technologies and create demonstration capabilities against select U.S. government use cases. These demonstration capabilities can show the government how:

- Users can save time and money with a more interoperable, consistent, and repeatable model.
- Products and systems can become more effective at meeting government needs and requirements.
- The AE can break down the data silos that multiple analytic tools typically create.
- Member companies can adopt consistent standards more quickly, saving them and the government time and money.
- The AE can develop a set of common standards and protocols for the analysis and analytics community.
- The AE can draw enriched products from the results of different companies' products.

2. Use case: Threat Assessment

2.1. User Need:

One use case selected by the Roundtable for collaboration in the AE is threat assessments. While threat assessments are a proven approach to lower the risk of targeted violence by proactively identifying, assessing and mitigating threats, the resource intensive methods utilized can make the process prohibitive in both time and money.

The U.S. Governments seeks analytic capabilities that reduce the costs of doing complex data exploration and increase the effectiveness in identifying trends and patterns of targeted violent events over time and space. These capabilities can help government analysts more effectively identify and assess threats in immediate time to control and neutralize specific threat actors.

2.2. Use case goal:

The goal for this use case effort is for select Roundtable companies to come together in the AE to demonstrate a more efficient and automated approach for conducting threat assessments using their respective and integrated commercial off the shelf technologies.

3. Threat Assessment Working Group

Centrifuge Systems, ESRI, Recorded Future and SAP NS2 are partnering in the AE to demonstrate how an integrated analytic system that efficiently applies commercially available solutions can more efficiently assess a threat posed by the militant Islam Group, Boko Haram in the Sub-Saharan African region.

SAP NS2 will employ its high speed data analytics platform (HANA) to ingest Armed Conflict Location and Event Data (ACLED), enrich it with threat intelligence from Recorded Future, add geospatial context from ESRI maps and then deliver results to Centrifuge Systems, which leverages its interactive analytics technology to enable threat analysts to more efficiently explore and share data and generate threat assessments.

4. Data set

The working group will use the open source Armed Conflict Location and Event Data Project (ACLED) data (http://www.acleddata.com/) as the key data source for its threat assessment effort (Clionadh, Linke, Hegre, & Karlsen, 2010).

5. System Architecture

Our multi-vendor system leverages integration to correlate and analyze multiple data types including geospatial, temporal, unstructured, and structured data, combined with utilizing visualization to understand the relationship between large quantities of events in near real-time.

ACLED data is used to link conflict actors to other conflict actors and the resultant number of fatalities within a specific country and within a defined timeframe. The data is stored in SAP HANA and visually analyzed in Centrifuge Analytics with Drill Charts, Timeline, Link Analysis, and Geospatial visualizations. Within Centrifuge Analytics, ESRI provides the map service and Geospatial Data Services for the Geospatial visualization of the conflicts based on the coordinates in the ACLED data. ESRI also provides visualizations in ESRI client applications.

Data enhancement is added by querying Recorded Future data around the dates of the event data contained in the ACLED data. The data is run through SAP's NLP engine with sentiment analysis. This data is then linked to the ACLED data that is visualized.

The visual data analysis presents event actors, dates, event types, country, and number of fatalities. With the enhanced data, the analyst is also able to see the opinions of the events, what people in the area fear, and, possibly, any Twitter intelligence leading up to the event.

The data is used to analyze two threat types: internal and external, based on the actors and their affiliation. Centrifuge Analytics is used to visually analyze the actors and affiliations to determine the type of threat. ESRI is used to visualize hot spots of activity geospatially.

6. System Demonstration

The system demonstrates how an analyst can perform temporal, geospatial, structured data and unstructured data analysis more efficiently using integrated commercial solutions.

As an example, the analyst is interested in looking at concentration of events that happened in all of Africa in the last two years. He/she zooms in on the timeline until only the last 2 years' worth of data are visible. This information is broadcast to the rest of the interface to filter out any events that did not occur within the selected time frame.

Now that the analysts have easily narrowed the time frame of events in view, they can go to the next step and focus on a specific geographical area. The analysts next filter geospatially by simply zooming in to the region around Nigeria, effectively filtering out any events that did not occur in that region.

Since the timeframe and region have been defined by the analysts they can now start to gain an understanding of what has actually happened in that region during that time period by looking at the tag cloud. The analysts can select a term in the tag cloud to show only events related to their selected term or they can enter a term of their own.

A MetaCarta service is included with each response. Using the Location-Finder from MetaCarta the user can retrieve more tactical information about the surrounding geographic area, it identifies geographic locations – public buildings, infrastructure, addresses, cities, land features, landmarks, etc. – and provides user with a catalog of the specific locations.

Each event is supplemented with a Recorded future query that analyzes and provides additional "Dark Web" content data not contained in the ACLED data set that identifies events leading up to a major event in 2015 referred to as the Fotokol Event which is a major Boko Haram attack on the town of Fotokol in Cameroon. The events signify an escalation of activity just prior to the event and battle damage assessment after the event. This data can be added to the ACLED data to give more insight on predicting future events via HANA predictive analytics.

With a subject matter selected the analysts can now switch the heat map to show the number of fatalities as opposed to simply the number of events within their selected region and time frame. From there they can zoom in on the map and select any event to view the structured data points that describe the event while leveraging the MetaCarta service to detect nearby features (e.g. schools, infrastructure, parks, notable landmarks/buildings).

7. Data Exchange Model

SAP ingests the ACLED data and adds a query for enrichment from Recorded Future. For each response, the user can access Recorded Future's retrievals from the dark web, specifically aligned with the ALED event type, and occurrence date. A link to the MetaCarta Location-Finder allows for enrichment adding local geo-features. The data is structured in SAP and provided to the Transfer service of the Data Exchange. ESRI also provides Geospatial data to the Exchange. A tool, such as Centrifuge Analytics will then extract and analyze the information from the Data Exchange.

8. System Flow

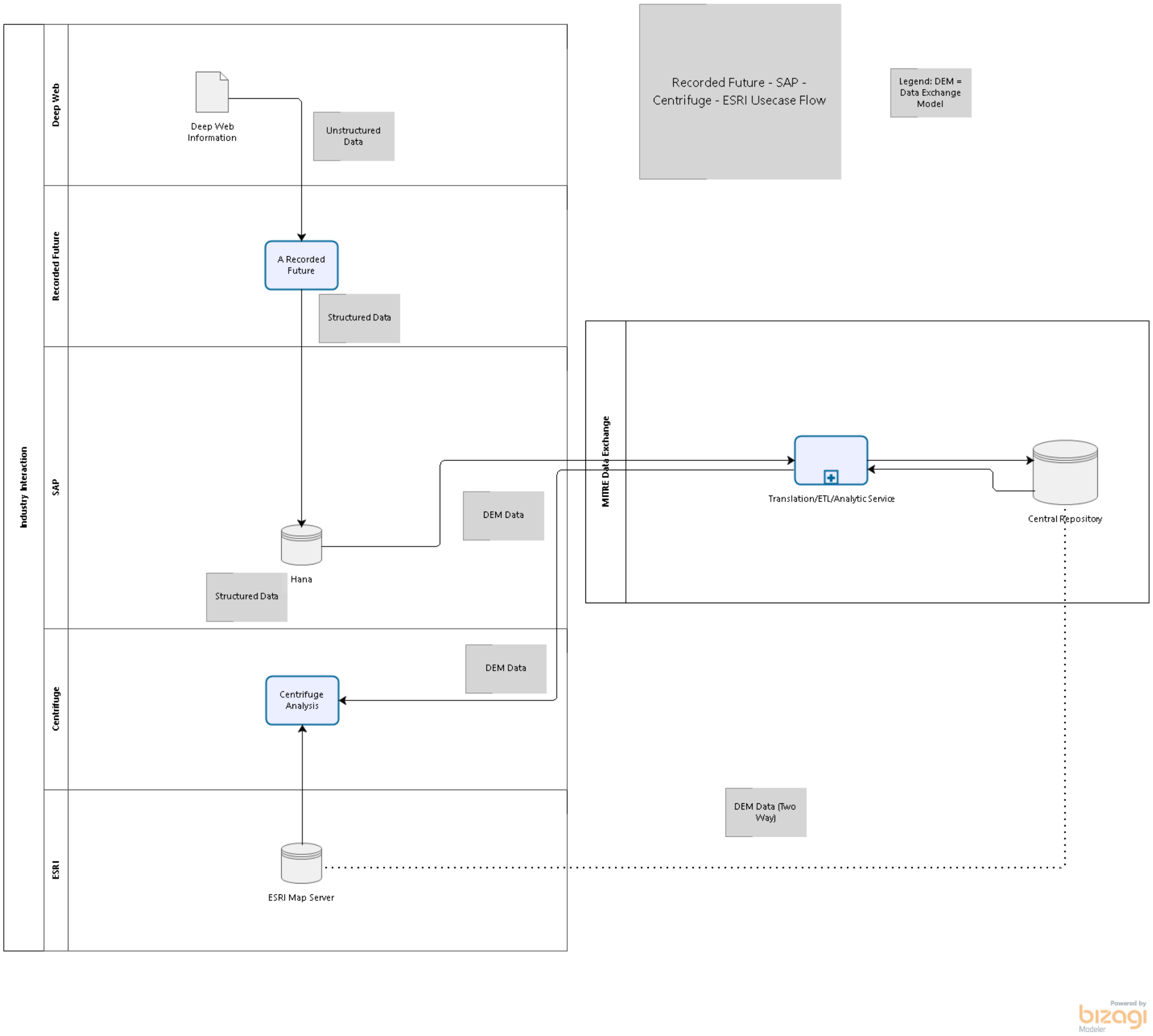
1. ACLED data set is ingested into **SAP/NS2 HANA**.
2. HANA provides analysis of ACLED's unstructured textual content (which some of the ACLED data entries contain) to produce sentiment analysis and entity extraction; this is performed within HANA and results will be put back into HANA.
3. Queries are added for users to selectively include **Recorded Future** "Dark Web" information linked to the structured and unstructured text of ACLED, retrieving further content that contains both unstructured and structured text to add more instances.
4. ACLED also provides geo-location information in the form of MetaCarta Location-Finder queries, and event locations which can be sent to **ESRI** for further enrichment.
5. These different sources, are brokered in HANA and are adapted into the Exchange's data model and moved into the Analysis Exchange (this is MITRE's principal responsibility).
6. Once in the Analysis Exchange, **Centrifuge** and **ESRI** access this to visualize the linked data information and map information.
7. Centrifuge expects to receive, at minimum, the following fields in JSON format from the exchange:

    - EVENT_ID_CNTY
    - ACTOR1
    - ACTOR2
    - ALLY_ACTOR_1
    - ALLYACTOR_2
    - EVENT_DATE
    - YEAR
    - EVENT_TYPE
    - COUNTRY

- LOATION
- LATITUDE
- LONGITUDE
- SOURCE
- FATALITIES
- **JSON Query designed for Recorded Future
- **GET-based URL to MetaCarta leveraging Latitude/Longitude from above

Note, in the pipeline above, several steps of enrichment are performed before any content reaches the Analysis Exchange (e.g., Recorded Future to HANA, HANA's own text analysis, and MetaCarta Location-Finder insertion). We receive the enriched results from HANA, which is what is adapted into the ontology. This is why we've limited these questions to just the connection points to the Analysis Exchange. There are still topics within the overall use-case that the team should address, such as what particular analytics are chosen, the orchestration between the analytics, and how the results will be represented in HANA.

9. Data Flow



10. References

Clionadh, R., Linke, A., Hegre, H., & Karlsen, J. (2010). Introducing ACLED-Armed Conflict Location and Event Data. *Journal of Peace Research, 47*(5), 651-660.