Analytic Technology Industry Roundtable Study:

Analytics and Use Cases

Dr. Adam Etches, IBM

Mr. Charlie Brown, IBM

Mr. John Stultz, SAS

November 2016

This page intentionally left blank.

Study TOR for Reference

Background

Over recent years the tools, techniques and procedures (TTPs) of organizations involved in intelligence gathering and understanding have become increasingly homogenized. With this alignment of TTPs across different domains and organizations it is now possible to generalize certain common generic use cases for both government organizations and commercial entities.

A use case in this context encompasses a general approach to a 'mission' or task centered operation. Specifics around data types and sources or collection of data will vary by organization, due to unique regulatory, industry, or legal considerations, but the foundational elements of the use case are the same.

Use cases are relevant to any organization, government or commercial, affected by criminal, unethical, and immoral activities, the aim being to maximize the value of the mass of information available to them and translate this into actionable intelligence for faster, more informed decision making.

Charter

This study will:

- Seek to define a set of the most common use cases across government and commercial organizations
- The use cases will be expressed as a set of commonly asked questions or requests for information (RFI's) and their expected outcome.
- The understanding of the use cases will be detailed in terms of the common TTPs usually employed.

Subsequent studies will leverage these use cases to evaluate the effectiveness in the government sector of COTS (common off the shelf software) that software vendors have developed for commercial entities. The key premise is that similar use cases and associated processes could be supported by the same foundational software.

Study Products

A catalog of use cases defined and peer reviewed, published to an open audience.

Introduction

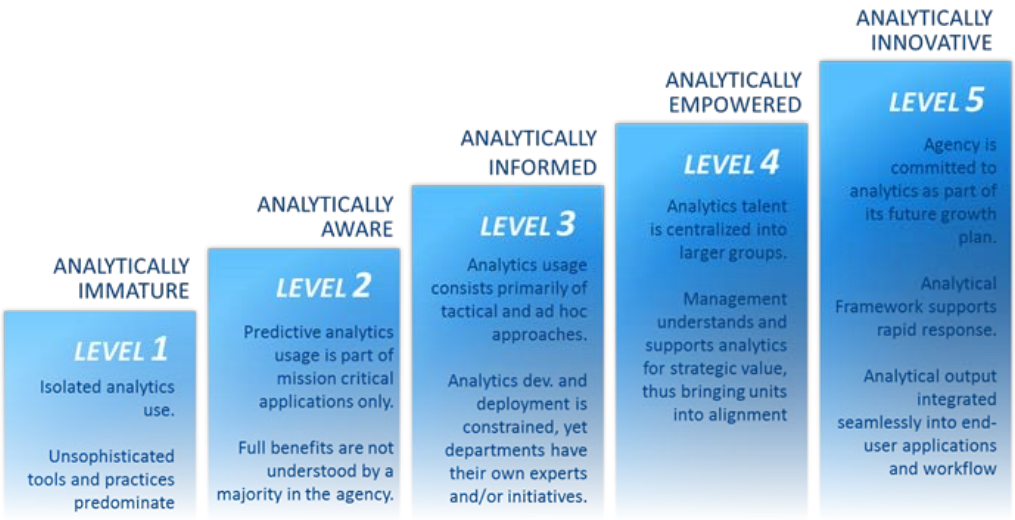
The core purpose of any analytic or investigative organization is to discern the insights needed for mission or task-centric activities: "*maximize the value of the mass of information available and translate this into actionable intelligence for faster, more informed decision making*"

This study group considered several approaches for describing use cases that are generally applicable across both government and civilian organizations. For example:

- Traditional processes and methodologies such as the OODA (Observe, Orient, Decide, Act) loop [reference: https://en.wikipedia.org/wiki/OODA_loop] , OBP (Object Based Production), and ABI (Activity Based Intelligence) [https://www.ncsi.com/diaid/2013/presentations/johnston.pdf]
- Source information used for missions and tasks, and how many organizations are moving away from traditional sources (high quality, lower volume, well-structured reports from trusted individuals or trusted systems) to ad-hoc and informal sources (variable quality, higher volume, variable structured, source and provenance often unknown). Whether we call this Big Data or Open data, the shift clearly impacts both the tools and the processes.
- Threat organizations and actors, which are shifting from traditional clearly defined threat organizations to informal or loosely organized groups: moving from symmetric to asymmetric threats.

These approaches all have merits, but in the view of the study team by themselves they do not provide an effective approach for grouping the techniques and procedures.

Instead, the study team is using the concept of maturity model (figure 1) or development level to describe how various techniques and procedures are typically combined to further the mission or task objectives.



When addressing a threat, an organization can only apply the tools and processes where it has the skills, capabilities and capacity to deploy. Otherwise, they expend more time and effort attempting (often unsuccessfully) to leverage these tools and techniques instead of focusing on the mission and task. The intent is not to overwhelm the analyst and analytical organization but instead empower them in their day to day operation.

The following table provides examples of a typical commercial pattern for each analytic development level. Note that these development levels are simply meant to provide a notional framework for review purposes and are not fixed or absolute: any single customer may exhibit traits that span across these development levels depending on their specific use case or industry.

Development Level	Example
1: Immature	Corporate security investigator given a physical security incident report and combines with other corporate information (building access logs, phone records, etc.) looking for obvious connections and contradiction
2: Aware	Insurance investigator charged with finding fraudulent claims or providers within medical claims given spreadsheets of customer claims and providers to look for fraudulent claims or providers with unusual patterns (such as different providers sharing the same address or fax phone number).

Development Level	Example
3: Informed	Fraud team investigates sales of counterfeit products or similar products to identify sources of the fraud, whether from authorized manufacturers and distributors or leaks of corporate intellectual property from within. Formal corporate product documentation is combined with fraud reports, open source intelligence, and employee communication (phone, e-mail), and when appropriate, personal financial information obtained in cooperation with law enforcement
4: Empowered	Hospitality corporation (casinos, hotels) maintains corporate wide view of customers, focusing both on identifying frequent customers (providing them a customized experience and rewards that encourages repeat business) as well as fraudulent customers with unusual patterns of activity (preventing the specific client from further fraud, and implementing pro-active measures to prevent others from committing similar fraud). For example, automated analytics could detect a spike in the redemption of points for cash-based of gift cards, and the combination of loyalty program information with employee shift information identifies a group of employees that register purchases from non-loyalty customers against their fraudulent rewards accounts (and then they redeem these points for cash gift cards).
5: Innovative	Financial institutions establish a cross-industry clearing house to share information on suspicious or clearly fraudulent international money transfers and the associated participants (sender, sending country bank and payment system, receiving county bank and payment system, etc.). This clearing house includes participation from law enforcement and responsible financial regulatory organization, as well as international counter terrorism teams. Secured access is implemented across the aggregated source information and intelligence, ensuring that participants can broadly share the appropriate anonymized information to identify specific incidents and patterns that are only visible when considering the end to end process across multiple participants. The non-anonymized information can be shared on a need to know basis only with those parties trusted by the source financial institutions – whether other financial institutions involved in the transaction, a clearing house task force that creates / shares black lists, or with the organizations that take the appropriate civil / criminal / anti-terrorism remedial actions

The remainder of this presentation will focus on two commercial use cases (3 / Informed and 5 / Innovative), describing the analytic tools and processes, and then compare each with an example within government. The commercial and government examples are described generically, and represent the combination of elements from commercial and government experiences.

Example 1: Counterfeit Product Project Office

A large clothing manufacturer faces financial and reputational risks due to counterfeit products available in large quantities. There is significant variation in the quality and fidelity of the counterfeit products, but some of them are of such high quality / high fidelity that counterfeiters appear to have access to the product design and specification details, and /or able to obtain parts or complete products from authorized manufacturers that in theory should be under sole source for the legitimate channels.

The manufacturer is receiving numerous reports and background information about the counterfeiting from a variety of sources (consumers, manufacturers, company employees, retailers, law enforcement, etc.) and does not have the organization, employee talent, or tools to effectively address the growing threat.

Therefore, the manufacturer decides to implement a single fraud investigation project office to identify the source of fraud. This project office is based in the supplier relationship organization, but will include representatives from all parts of the manufacturer.

Since time is of the essence, the project office quickly submits an RFP looking for a turn-key solution that can be rapidly deployed. Key elements include:

- Ability to connect to and load information from key corporate data sources. Since this team will be dealing with sensitive information (shipping information, corporate IP, employee records, supplier information, customer, and acquired evidence for law enforcement) the decision is made to create a separate repository for this team instead of using existing corporate IT systems
- Tools for exploring open source (aka internet) for information related to the counterfeiting operations, whether advertisements for the counterfeit goods, information about the counterfeiters (who often collaborate / discuss

the operations on social media) as well as details on legitimate suppliers who may have hidden business operations related to the counterfeiting

- Support a small but active set of investigators (initially 10, but slated to grow to 50). These investigators will be working independent cases, but they should be able to search for relationships between the individual cases in order to identify larger counterfeit networks.
- Case management tools that can be used as a lead (tip) management system, effectively prioritize and assign these leads, provide an adaptable reporting (dashboard) tool, enable documentation and dissemination of adjudicated cases (both internally and externally, for example to law enforcement), and where appropriate allow investigators to create new leads when they discover additional counterfeit examples during the investigation.
- Specific analyst facing tools, enabling them to explore source information, identify key evidence ("intelligence") and combine this evidence to identify counterfeit actors / organizations ("insight").

The project office assembled a team of potential investigators to gather input on the required analyst tools and found a wide range of recommended tools due to their diverse backgrounds -- law enforcement, financial, market analysis, supply chain, etc. These tools were most frequently mentioned, and were included as required capabilities in the RFP:

- Unstructured Text Analytics: extracting key information from unstructured documents and web sites, converting into a structured format that could be consumed by other tools. OCR whether from images, faxes / scans, or PDFs is highly desired since many tips are in these formats
- Entity Analytics: examining properties about individuals, organizations, and activities, and discerning relationships that are not obvious or hidden (for example, the phone number for delivery pickups used by a counterfeit organization is the same phone number used by a legitimate supplier)
- Link / Network Analysis and Visualization: exploring the objects and relationships between the source intelligence and looking for common patterns and connections (for example, trying to discern who is the controlling the counterfeit organization versus who is an operative)
- Geospatial Analysis: attempting to identify individuals and activities who work in close proximity (for example, in a large urban area the counterfeit organization may be in a building that is around the corner from the legitimate supplier).

- Temporal Analysis: identify a sequence of events on a timeline to track the spread of information or products (whether leaked corporate IP or the distribution of a batch of counterfeit goods).
- Pattern Recognition / Statistical Analysis: tools that can examine a large set of information (such as company phone records) and look for identifiable features or trends that could indicate significance (for example, how does the calling pattern of insider threat differ from normal calling patterns).

With the RFP complete, the manufacturer puts the proposal out for competitive bid, and receives multiple responses from industry that in general fell into these two categories:

- Custom integration contract – using off the shelf products from multiple suppliers combined with an 18-month development effort to provide the needed capabilities. High cost, long lead time, richest function / best of breed.
- Single product – using a single off the shelf product that provided most but not all of the capabilities, combined with limited services / integration. Most were deficient in one or more areas, such as Case management, data acquisition. Lowest cost, lowest lead time, least function

The manufacturer decided on a middle ground and asked for these vendors to revise their bids to provide an immediate solution that provided partial capability upfront (built on tools from one or two vendors that had out of the box integration) and then provide additional capabilities in a subsequent phase with targeted integration to point tools that provide the additional capabilities.

The winning updated proposal provided most of the capabilities for data-centric acquisition and investigator centric analytic / visualization within three tools (data platform, analysis / visualization, and geospatial), and deferred the core case and lead management system to a subsequent phase.

The manufacturer had an existing case and lead management system that they could leverage for this project, although this would lack the desired tight integration. Instead of having a unified “single pane of glass”, the investigators would have to switch between the two application spaces.

The above description was made in the context of a commercial organization, but the essential elements are similar to proposals within government opportunities, in particular in civilian and social services departments where they are

typically not looking to procure or create a single enterprise platform. Yes, compliance with key government IT standards are critical; however, the procurement is a point acquisition (obtaining a solution for a specific purpose).

This paper does not intend to imply that the civilian and social services departments lack investigative / analytic maturity and experience: in fact, often just the opposite is true in particular in well-established programs and organizations that have been dealing with fraud for decades. However, the primary focus in these organizations typically are focused on the individual mission and time to value in enhancing that mission, and that tends to focus procurement requests for analytic / investigative solutions versus analytic / investigative capabilities or platforms.

Example 2: Financial Crimes Fusion Center

A large financial institution with a global footprint is going through a risk mitigation modernization effort in order to meet new regulatory compliance requirements and further reduce operational and financial risk. They view risk mitigation as being an imperative that can only be achieved at the industry level versus an individual corporation, so they reached out to several competitors and created a financial crimes consortium. The initial consortium conference resulted in the following working groups:

- Regulatory compliance
- Data governance
- Network security risk (internal/external threat)
- Money laundering & terror finance prevention

The purpose of each group is to identify the strengths, weaknesses, as well as external opportunities and threats in each respective area. This includes identifying existing and potential external stakeholders who would benefit from having access to the financial institutions data on a need-to-know basis. The entry point for a stakeholder would be enabled by a self-service “clearing house” portal with role-based security supporting data access specific to the needs of the user.

External stakeholders include law enforcement, partner financial institutions, counter terrorism teams and others on a need-to-know basis. When appropriate, data will be aggregated and or anonymized.

Some key elements and possible scenarios that the working groups will address are:

Regulatory Compliance

- Integrate business policies and procedures to ensure regulatory compliance for Customer Due Diligence, Anti-Money Laundering controls, ISO 27001 standards for data and information systems and individuals, ISAE 3000 assurance for business processes as well as other compliance functions (Dodd-Frank, FATCA etc.).

Data Governance

- Support external stakeholder with appropriate access control templates for accessing information within a "self-service clearinghouse" portal.
- Support data lineage capabilities, real-time data monitoring, dashboards and scorecards to check and control data integrity over time, monitor change data capture, include data stewardship capabilities (tracking and monitoring data sourcing, data transformation, cleansing and storage) and support web-based dashboarding and business rule exception monitoring for reporting and remediation.

Network Security Risk (internal/external threat)

- Identify areas for improving the monitoring, sharing, and responding to cyber threats and vulnerabilities.
    - Identify administrative and operational gaps that currently impede data integration of network security data and financial operational data. For example, phishing scams often enable an outside entity access to internal systems and the exploitation of those systems might be timed in a way that enables other financial operational exploits via wire transfers, money-laundering or other sophisticated financial crimes to take place. While technology can not totally compensate for human weaknesses when it comes to phishing scam vulnerabilities, the ability to baseline network behaviors and enable early detection of internal network threats can mitigate advanced persistent threats from taking hold within a system. Through the ability to cross-reference network risk data with operational risk data that is often in disparate siloes, vulnerability assessment analysis can have a significant impact on reducing financial loss.

Money Laundering & Terror Finance Prevention

- Assess current policies and procedures that are in place to detect and mitigate money laundering and financial crimes. This includes further development for a cross-industry clearing house to share information on suspicious or clearly fraudulent international money transfers and the associated participants (sender, sending country bank and payment system, receiving county bank and payment system, etc.). An examination of controls for Customer Due Diligence and an assessment of security protocols around innovations in new payment methods (NPMs) and digital currencies will also be performed. The systems and controls would need to enable the following role-based user scenarios:
    - A Department of Homeland Security (DHS) Investigations Special Agent has received information from an informant that a commercial company in Fort Lauderdale is importing over-valued gold scrap from the Dominican Republic. The agent has a need to check all available data bases to determine what additional information on the commercial company's financial transactions might be available.
    - An Assistant U.S. Attorney in Minneapolis suspects that local unlicensed money remitters are involved with an underground hawala financial network that transfers money and value to the Horn of Africa via Dubai. He directs the regional Suspicious Activity Report (SAR) review team to examine financial and other data to determine if sufficient information exists to pro-actively initiate an investigation.
    - An Alcohol, Tobacco, Firearms and Explosives (ATF) manager in Washington D.C. is concerned about a recent spike in human source reporting in El Paso that indicates an increase in the trade of narcotics for weapons. Narcotics come north across the Mexican border and weapons go south. In pondering various investigative strategies and appropriate resources to deploy, the manager needs to first direct an analyst to query multi-source databases (including financial transactions) to see if other regions in the country are also experiencing similar trafficking patterns.
    - Local law enforcement and intelligence professionals need access to data that will help them make cases. Within the clearing house there must be a mechanism for secure access and intuitive manipulation and interpretation mechanisms so they can query and apply filters to meet their unique mission requirements.

The above scenarios are made in the context of a financial crimes fusion center, but the essential elements are similar to proposals within government opportunities where there are multiple stakeholders who need to access information relevant to their mission and share results. In particular, agencies with different groups / mission requirements who need to access information and generate analytic insights on their own and integrate data-driven insights in real-time.

Similar to the Counterfeit Product Project Office use case, Department of Defense, Intelligence Community fusion centers are typically not looking to procure or create a single enterprise platform. While they too have compliance requirements with key government IT standards, the procurement is a point acquisition (obtaining a solution for a specific purpose).

Again, a key element is organizational: the culture and organization in these departments is often an investigative / analysis culture because they ***are*** supporting law enforcement and defense intelligence. These organizations have a strong bottom-up orientation that drives the needs of this specific community and puts a greater emphasis on data preparation and role-based information dissemination to support an "Object Based Production" environment.

Analytic Tools Considerations

The case studies above provided a narrative overview of two analytic or investigative use cases, whereas the table below provides a more explicit treatment of the common characteristics and how they could vary between the Analytically Informed and Analytically Innovative development levels.

Example	Level 3: Analytically Informed	Level 5: Analytically Innovative
Use Case	Counterfeit Products	Public / Private Industry Fusion Center
Use Case Description	Fraud team investigates sales of counterfeit products or similar products to identify sources of the fraud, whether from authorized manufacturers and distributor or leaks of corporate intellectual property from within	Financial institutions establish a cross-industry clearing house to share information on suspicious or clearly fraudulent international money transfers and the associated participants (sender, sending country bank and payment system, receiving county bank and payment system, etc.).
Organization	Department alignment	Inter-Enterprise
Acquisition Objective	Obtain turnkey solution to meet immediate business needs (“greenfield”)	Obtain capabilities that can be combined with existing or home grown capabilities to create tailored solution to meet long term business needs (“brownfield”)
Procurement	Although used across the enterprise, the solution is typically procured by a single line of business organization, and over the time the usage is extended to other parts of the business	The capabilities are procured at the overall business level by the IT organizations, and then integrated into the common platform / solution that is made available to all lines of business.
Data Sources	external information sources / no additional insights	Shared data across multiple enterprises

Example	Level 3: Analytically Informed	Level 5: Analytically Innovative
Number of Data Sources	>100	>1000
Cumulative Size of Data	>100GB	>5TB
Data Acquisition and Curation	Automated tools in place to acquire new or updated data sources, but analysts are typically required to curate and extract new intelligence. "Analyst spend 80% of their time curating data and 20% of their time exploiting the resulting intelligence"	Automated tools in place to acquire new or updated data sources, as well as to curate, extract new intelligence, and correlate with existing intelligence. An example of this type of process is Object Based Production. "Analysts spend 20% of their time curating data and 80% of their time exploiting the resulting intelligence"
Intelligence Repositories	Accumulated intelligence per department	Intelligence shared across enterprises
Schemas	Shared at department level	Cross enterprise agreement on schema
Number of Users	25 – 50	100 +
Size of Collaboration Analytic Teams	1 – 5	10 +

Example	Level 3: Analytically Informed	Level 5: Analytically Innovative
Collaboration Considerations	Case driven: Users are typically organized by line of business and work a single investigation end to end, reaching out to other users in the same of different lines of business when overlap is detected between cases.	Expertise / Skill drive: Users are typically organized by their expertise – whether business / domain or tool / analytic – and work in teams on many cases.
Case Trigger	• External incident report • Analyst discovered pattern / event	• Analytic discovered pattern / event • Analyst discovered pattern / event • External incident report
Case Output / Usage	• Adjudication of network of events • Reports to share Insights	• Adjudication of network of events • Real-time prevention • Immediate communication to broader community • Reports to share insights
Analytical Approaches	• Visual Analysis • Pattern Recognition • Entity Analytics	• Visual Analysis • Pattern Analysis • Entity Analytics • Predictive Analytics Machine Learning
Analytical Timeline	• Re-active, driven manually be the user when investigating a case • Deferred, when data is stored (typically run batch or scheduled)	• Immediately, on data acquisition (Real time / Streaming) • Deferred, when data is stored (typically run batch or scheduled) Re-active, driven manually be the user when investigating a case

In Summary

This study has sought to define a set of common use cases that are relevant across government and commercial organizations for the purpose of discerning the insights needed for mission or task-centric analytical activities. We suggested that the essential elements of an analytic or investigative solution are effectively the same, whether in a government or commercial environment. In order for the reader to gain a perspective of where insights are relevant, a general 5-level Analytic Maturity framework is used to support the differentiation and perspective of where one might be and what it looks like to move up or down in analytic maturity. This is important since this often becomes the perspective for understanding how assets are currently allocated and prioritized and what barriers might need to be overcome to reach a higher level of analytic adoption. The use cases and tables highlight how, for example in requests for information (RFI's), a desire for a higher degree of analytical capability can map to a path to adoption. Ultimately the hope is that this study will provide a context for how to map desired outcomes with an agency or organization's level of readiness**.**