NATIONAL GEOSPATIAL-INTELLIGENCE AGENCY

Know the Earth... Show the Way... Understand the World

2020 ANALYSIS TECHNOLOGY PLAN

DEFINING THE ANALYSIS TECHNOLOGY VISION FOR 2020

1. Introduction: Future State Vision and the Role of Analysis Technology

NGA's Future State Vision is of an agency that is both an indispensable geospatial intelligence (GEOINT) provider and an enabler of multi-INT integration for the Intelligence Community (IC). The Future State Vision calls upon us to transition from a provider of static products to a provider of dynamic GEOINT content, analysis, and services, and it requires us to make our analysis structured, metadata-tagged, and accessible. This must be accompanied by an evolution of our tradecraft to fully incorporate big and open data solutions, advanced phenomenologies, and advanced analytics that more thoroughly leverage the geospatial content inherent in multi-INT data sources.

Achieving this Future State Vision is equally dependent upon the technology and tradecraft of the Analysis Directorate. Technology must enable these transformations, while tradecraft must operationalize them. This will involve a concerted integration and evolution of technology and tradecraft to support the Structured Observation Management and Activity Based Intelligence initiatives while leveraging the analytical, production, and collaboration opportunities presented by the other initiatives (Map of the World, Next-Generation Collection, The Globe, Customer Engagement, and Unified Data Strategy).

2. Vision: GEOINT Exploitation and Analysis

With increasing amounts of geospatial and imagery data, a new paradigm is needed for GEOINT exploitation and analytics. GEOINT analysts need the knowledge and capabilities that allow them to spend less time exploiting GEOINT primary sources and more time analyzing and understanding the activities, relationships, and patterns discovered from these sources. This paradigm shift must also include more efficient and dynamic ways of publishing data, knowledge, and intelligence gleaned from the exploitation and analysis of these sources.

To achieve this, analysts need a unified GEOINT platform that aligns disparate tools, algorithms and capabilities into an interoperable, data-centric exploitation and analytical system of systems – an 'integrated analytic environment.' This environment will enable analysts to discover, access, organize, display, exploit, analyze, and expose data and intelligence

within a single, unified framework. As a data-centric system, the integrated analytic environment will natively display any geospatial, imagery, or intelligence data necessary for GEOINT analysis, and integrate that data at the point of analytical use. Analytic capabilities, to include exploitation tools, are driven by data in-use and provided a la carte within this unified environment.

The integrated analytic environment must also leverage opportunities presented in the automation of exploitation capabilities. Automation may be in the form of tools and algorithms that reduce data dimensionality, provide investigative cues based on data correlation and object/change detection, and help the analyst transition from a forensic-based to a model-based approach to GEOINT analysis.

Analysis technology must be deliberately evolved to push the bounds of the possible, both through more automated and flexible tools and a co-evolving tradecraft that fully leverages and demands more from these tools. There are four categories of technologies that Analysis must provide a vision for, and subsequently provide the requirements: Research & Discover, Access & Visualize, Exploit & Analyze, and Expose & Report. The capabilities driven by Analysis in these categories are not stand-alone; rather, they intersect, drive, and feed from content services managed by Source, customer services and business analytics managed by Xperience, data standards managed by Office of the Chief Information Officer (OCIO), an information architecture managed by IT Services and OCIO, a workforce hired and developed (trained and educated) by Human Development, and Research, Development, Testing & Evaluation (RDT&E) performed and managed by InnoVision. The Analysis capabilities presented are each a system – a Discovery system, a Visualization system, etc. – but these systems must be linked via common, standards-based interfaces into a system of systems to enable interoperability across and through each category of capabilities, across network domains and regardless of the analysts' physical location.

Research & Discover

Key technology needs:

- *Centralized cross-domain, data-agnostic federated search capability*
- *Natural language/semantic query capability*
- *Automated metadata tagging for all archival and current GEOINT data holdings to specify spatio-temporal and intelligence/domain context*
- *'Deep learning' algorithms for automated data discovery*
- *IC desktop dashboard with real-time, contextually-relevant content streams*

Broader and deeper analytics are reliant upon analyst ability to find data and content relevant to their intelligence domain and intelligence questions. GEOINT analysts will be able to find any and all data relevant to their intelligence question, and this data discovery will be immediate. All NGA data will be spatially, temporally and contextually indexed and service-enabled. NGA will employ an index aggregator and federated cross-domain search that allows for simultaneous multi-database complex queries, and a semantic processing engine that understands and translates common phrasing to machine-readable query language.

Data discovery will be operationalized by a central query capability that rides on top of the analysts' exploitation and business environment. The query capability will execute across security domains, providing not only NGANet/JWICS-present data and products, but also SIPRNet, NIPRNet, and open internet. Contextually relevant results will be conveyed to the investigating analyst up to the appropriate security level, and the query capability will enable viewing/visualization of the data in the integrated analytic environment.

Understanding where and what data are available to analysts is facilitated by automated metadata tagging of NGA- and NSG/ASG-produced and consumed data. All data, phenomenologies and intelligence products will be automatically tied to spatio-temporal and

intelligence context. Un-queried data will also find its way to analysts through automation and 'deep-learning' techniques: big data algorithms will constantly process NGA and IC data holdings for phenomenological, spatial, temporal and relationship signatures for deviations from normal, and provide analysts cues/ indicators for further exploratory analysis.

Analysts in 2020 will be supported by an IC dashboard that serves as a real-time content feed for relevant intelligence data. The IC dashboard will highlight issue-specific intelligence since the user's last login (e.g., overnight intelligence data from forward-deployed analysts or ASG partners), and this intelligence will be seamlessly visualized and exploitable in the analysts' integrated analytic environment.

Access & Visualize

Key technology needs:

- *Aggregation of data access points into a unified/common distribution framework*
- *Unification of content visualization into a single, integrated analytic environment, rather than disparate platforms*
- *Open, standards-based data models for interoperability*
- *Service enablement of all NGA data sources*
- *Immersive visualization at the analyst's desk*
- *Mobile exploitation and analysis capabilities*

As the Agency transitions to the Intelligence Community Information Technology Enterprise (IC ITE), GEOINT analysts will have the potential to access a broader array of traditional and non-traditional geospatially referenced data sources. Critical to leveraging the access opportunity effectively is streamlining the framework through which data is accessed. By 2020, the majority of data will be accessed through a unified/ common distribution framework, rather than today's approach of separate access points for individual data sources.

As data access points are unified, the visualization of these data sources must also be unified. The integrated analytic environment will integrate the visualization capabilities of currently disparate platforms and datasets. Data fusion will evolve from a preferred analytic method to a natural state of data: all spatially referenced data will overlay in an integrated analytic environment, and data/content will be spatially accessible. For example, an analyst exploiting an image will have immediate in-image access to an available FMV feed (or hyperspectral cube) temporally consistent with the image they are viewing and the intelligence question they are investigating.

Access to datasets will be through spatial awareness of an integrated analytic environment. Analysts will access data without leaving the analytic environment, such that complementary data sources will be available within their spatial view. For example, an analyst investigating an intelligence question will be able to effortlessly call up alternate available phenomenological views from within the analytic environment. Access will also be perpetual: analysts will have the option to enable real-time source updates within the analytic environment.

Data and tool interoperability is an essential link to allow Analysis, and the enterprise, to keep pace with technological change and rapidly integrate new (or replace old) capabilities and datasets into the analytic environment. All future GEOINT data models will be based on open standards. This will facilitate true data and intelligence integration by removing proprietary barriers to content access, exploitation and analysis.

All data stores utilized and populated by the Analysis Directorate will be service enabled using open standards-based protocols. Service enablement will ensure Analysis-produced and consumed data are accessible by our intelligence partners, and can be shared through customer-specific geospatial data viewers and application programming interfaces (APIs). Service enablement also allows tools/algorithms within the integrated analytic environment to readily access, combine, and correlate across multiple datasets, and will facilitate the integration of intelligence content within Map of the World-based views.

Data visualization in 2020 will be a diverse, user-driven experience that drives exploitation and fuels analytic curiosity. Powered by Map of the World, analysts will

be able to visualize content through intelligence views, within the integrated analytic environment, that best model the question at hand and the analyst's analytical approach. For example, common GEOINT datasets may be rendered and layered differently for unique mission-sets, and unique intelligence questions within a mission set will drive unique content visualization.

New paradigms in data collection, from basic steps such as collecting structured observations to more advanced relationship and activity models, will open the door for new ways for analysts to visualize the content available to them for GEOINT analysis. Three-dimensional viewing will be commonplace, but spatial dimensionality will be expanded beyond our currently limiting XY and Z construct. Visualization of 'data space' will provide added insight and context on an activity or entity, for pattern discovery and definition, and for anomaly detection. Understanding, and visualizing, the proximity of object attributes, activity times, and entity relationships in data space expands our ability to model the behaviors of our adversaries and provide anticipatory intelligence.

New visualization capabilities will enable teams to work together in a multi-user visualization environment where they occupy the same data-space and landscape, but with unique perspectives (e.g., multi-player video games). Each team member will be able to collect and analyze observations, transform and analyze imagery and other phenomenologies, communicate with team members inside the exploitation environment, and visualize/utilize the results of each other's analytical work (as desired). This will support the evolution of the agency workforce structure to virtually integrated analytical groups, but must at the same time control for varying security requirements and access constraints. Visualization of the integrated analytic environment will be equally at home in large collaboration spaces: as mobility becomes a regular component of operations, the desktop analytical environment will seamlessly transition to collaborative environments, allowing multiple analysts to connect, interact and exploit in tandem.

Exploit & Analyze

Key technology needs:

- *Automated algorithms for object (entity and activity) detection, vectorizing, and attribution*
- *Automated algorithms for assigning and validating graph-based relationships between objects*
- *Enterprise-level database(s) for structured observation capture, sharing, and analysis*
- *Decoupling of technical capabilities/algorithms from stove-piped analytical platforms*
- *Touch-, gesture- and voice-based interaction within the integrated analytic environment*
- *Enterprise-level graph databases*
- *Analytic modeling framework for centralized GEOINT model management, sharing, and testing*

Exploitation in 2020 will be a joint effort between humans and computers. Expanding data volumes – the influx of big data, the 'internet of things,' and location awareness of nearly all forms of technology – will provide a wealth of sources upon which NGA will exploit for geospatial intelligence. This will require reliance on automation to capture and characterize observations and activities, with reliability estimates at the attribute level, and to assemble related data into analyzable networks. Cognitive computing and neural networks, through natural language processing and machine learning, will help us both capture structured observations and make sense of the structured and unstructured data volumes at our disposal. The role of human analysts will evolve over the next six years as computers assume an active (vice today's mainly passive) role in GEOINT exploitation: human analysts will spend more time interrogating and validating the machine's work, exploiting the gaps and applying broader contextual awareness to the algorithmic outputs.

As with visualization, exploitation and analysis will be user-driven experiences that are unique at multiple

user levels: the individual, intelligence question, mission set, place in the analytical model, and place in the analyst workflow will all drive the experience presented to the analyst within their integrated analytic environment. Exploitation and analysis will also be data-driven: as different datasets require different exploitation capabilities (e.g., open-source video vs. multispectral imagery), the exploitation environment will present analysts with capabilities relevant to their active data and unique user condition.

This user-driven approach to exploitation and analysis will be enabled by a decoupling of analytic capabilities from traditional analytic software suites and packages. Rather than needing to load imagery into an electronic light table (ELT) for histogram adjustment or applying an algorithm, these capabilities will be available in the integrated analytic environment as a la carte, on-demand functionality.

Capturing structured observations, relationships, and judgments will be an analytical habit, as data models will exist for all intelligence questions. This will be supported by enterprise-level relational, non-relational, and graph databases that are seamlessly linked, traversable in any direction, and spatially discoverable and accessible.

Central to exploitation and analysis is the ability to interact with and manipulate intelligence data to discover, identify, and characterize activities and entities. In 2020, interaction with the integrated analytic environment (and business environment) will predominantly be in the form of touch, gesture and voice. Keyboard and mouse interaction will be significantly phased down over the next six years in favor of these more natural, intuitive and flexible modes of computer interaction. Precision controls that enable accurate mensuration and targeting will augment motion-based data browsing, manipulation and exploration, and touch- and voice-based data capture, exploitation and analysis.

GEOINT analysis in 2020 will rely heavily on the use of analytic models to discover, validate and anticipate activities and events. A modeling framework will be implemented that centrally manages and executes our GEOINT analytical models, and this will be accessible through and integrated with the integrated analytic environment. Some models will be continuously evaluated against available data, allowing analysts to test multiple hypotheses and become more exploratory in their analytical operations. Other models will be executed independently by analysts against data subsets.

A model-driven approach to analysis requires the automation of many of the analytical tasks performed by analysts today. Models will know when a certain piece of derivative data is needed for correlation or validation – such as a view shed, least cost path or principle component analysis – and models will create or access these derivative datasets using prescribed mathematical methods. Thus, the modeling framework will connect to and leverage the analysts' spatial, imagery and statistical analysis tools within its own (server-side) analytical sandbox. The modeling framework will also be directly tied into the Source directorate's collection tasking framework, enabling model-driven automated tasking of GEOINT sources and other Intelligence sources for a specified time-space window. The use of model automation will expand the opportunities for analysts to perform exploratory research and analysis to improve and refine the analytic models, and to discover intelligence only ascertainable through the lens of human context.

'Deep learning' and artificial neural networks will play an important role in GEOINT analysis in 2020. These artificial intelligence capabilities will help us identify and refine our models by parsing and correlating the voluminous data streams at our disposal. As big data becomes an everyday reality for GEOINT analysis, big data algorithms will enable analysts to quantify and characterize 'normal', and detect spatial behavioral anomalies – that is, 'abnormal' – that serve as indicators for future significant activities or events. Approaching big data analytics with the appropriate questions will yield results that allow inference, analytic model refinement, and anticipatory intelligence. Use and correlation of multiple geo-referenced data and intelligence streams will enable analysts to more clearly distinguish noise and random outliers from significant anomalies that require deeper investigation, collection tasking, intelligence reporting, or model adjustment.

With the help of 'deep learning' – technology building upon artificial intelligence capabilities – we will develop deeper understanding of the data interrelationships

when modeling spatial behaviors and activities. With this deeper understanding, analysts in 2020 will begin to expand beyond model-based hypothesis testing and into spatial behavioral and activity/event simulations through spatial agent-based modeling. Spatial simulations will allow analysts to delve deeper into what-if scenarios, test the relationship between multiple hypotheses, and explore/investigate the genesis of activities as an event evolves over time.

Analysts will have at their disposal, seamlessly integrated within the integrated analytic environment, analytical tools appropriate for any data type being analyzed: imagery, non-imagery phenomenologies, video/motion imagery, geospatial, statistical, and graph. No longer will the analyst have to utilize unique, stove-piped applications for content-specific analytics, nor will the analyst have to convert data to application-specific (or be stuck with proprietary) formats for analysis. Through open data principles and governance, the integrated analytic environment will manage and perform data/content transformations for analysis, a process which is invisible to the end-user.

Publish & Report

Key technology needs:

- *Data-centric publishing platform that exposes GEOINT observations and judgments to customers in near real time*
- *Dissemination and messaging services that seamlessly integrate with the integrated analytic environment, allowing for data publishing and alerting without requiring reformatting, exporting, or porting*
- *Interoperable structured data visualization capabilities*
- *Real-time automated report generation capabilities from structured data*

The scope of NGA products will expand in the future. While formal products will remain a trusted means of communicating longer-range intelligence analysis, source-specific exploitation results (structured observations: imagery intelligence and geospatial data) and analytic judgments will be published to the IC through streaming data in near-real-time. These data feeds will provide not only the content, but also the quality and veracity of each observation and judgment, ensuring that intelligence reaches our customers quickly while also providing insight into the level of vetting data has been subjected to. Accomplishing this requires the implementation of a data-centric publishing service that streamlines the movement of knowledge from analyst to customer, allowing intelligence content to be accessible to our mission partners and customers, and internally, with minimal lag.

This publishing service and the supporting tools will be an inherent component of the future integrated analytic environment, allowing analysts and content managers to publish their intelligence findings as both human- and machine-consumable data. The publishing service will be connected to a messaging service that allows analysts, whether NGA, NSG, ASG, or IC, to subscribe to and be notified about GEOINT content associated with a known object, entity, activity, or space and time. Data publishing will be fully integrated into the IC's Object-Based Production (OBP) framework, allowing for cloud-based intelligence centralization driven by IC-recognized objects. Published data, however, will exceed our support of the current implementation of OBP: complex stand-alone visualizations, curated GEOINT feeds, and anticipatory model outputs will fall within the domain of GEOINT reporting in the 2020 timeframe.

NGA formal products will also transition from static to dynamic. Leveraging capabilities presented by Xperience's Globe and Source's Map of the World, customers will no longer need to spend time searching through as-built, single-use products for our GEOINT-derived knowledge: dynamic data-centric intelligence views will be the norm. The temporal dimension will grow in importance, not just in data analysis, but in product development and contextualization. New methods of data interaction, visualization, consumption, and integration will increase our ability to show the changing intelligence picture over time, along with the spatial and temporal evidence leading to analytic judgments. Partners and analysts, both internal and external, will have the ability to consume the dynamic product as well as the data and sources tied to that product for their own intelligence workflows and models.

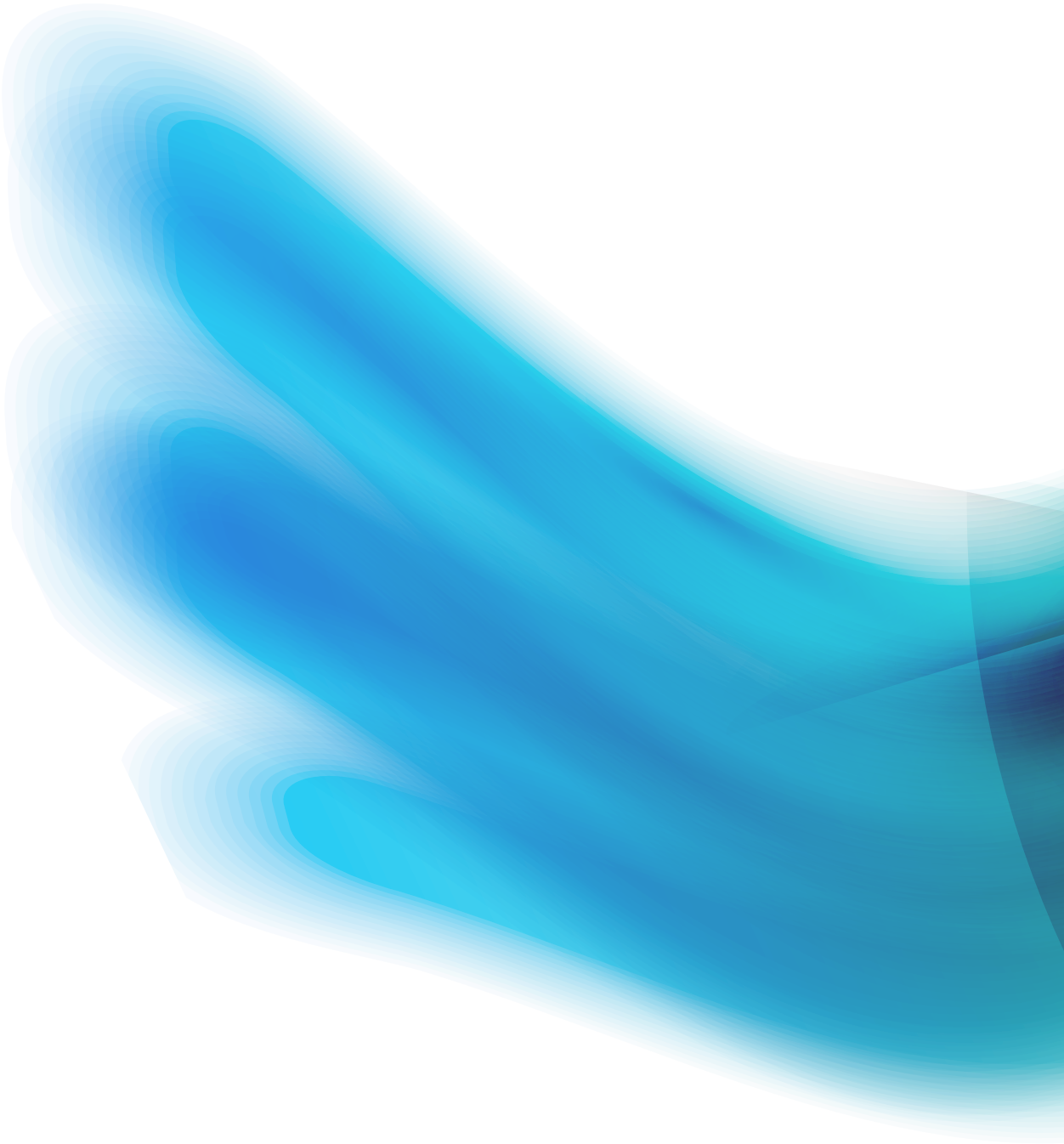
3. Conclusion

Technology will offer a wealth of possibilities for advancing NGA's analytic capabilities. NGA will soon be able to leverage the elastic compute resources of IC ITE for processing increasingly larger volumes of data, and machine learning and automation will help to break through resource-intensive exploitation bottlenecks. Breakthroughs in the ability to collaboratively work and interact with data will lead to an entirely new environment for performing exploitation and analysis.

The workforce is the most important variable in the future technology equation. Operationalizing new technology will require substantial amounts of training and new skillsets among the analytic cadre. The workforce must also be empowered to leverage these technologies against their key intelligence questions. This empowerment requires a cultural shift away from our traditional model of intelligence production through single-source exploitation to a new data-centric model. This new model will rely more heavily on the evaluation, analysis, and synthesis of a multitude of streaming sources of data, and the dissemination of intelligence content at multiple points within analytic workflows.

NGA must build a technology environment that attracts the best and brightest analytical and scientific minds the nation has to offer, sufficiently competing with both the largest technology firms and the innovative startups. Not only must NGA provide an environment where leading edge geospatial and analytic technology is implemented within our doors first (ahead of the commercial world) to address our most pressing national security threats, but simultaneously we must also provide opportunities for innovation to flourish among our workforce. Technology cannot accomplish this alone, however: adjustments in information, IT, and acquisition policy are required to make these changes.

Operating on the leading edge of technology requires some acceptance of risk, but the reward is a GEOINT agency that defines the direction of the discipline rather than responds to it. It provides our analysts the opportunities to push technical boundaries rather than be encumbered by them. Additionally, the risks will be mitigated through the commercialization and commoditization of geospatial data, such that the rest of the IC will be rich with geospatial data from commercial and open sources. Because of this proliferation of geospatial data, it is ever more important for NGA to re-envision its value proposition to the Intelligence Community as one that is not limited to being a provider of authoritative data and base maps. Rather, through broader and deeper analytic capabilities, achievable with the help of technology that automates the routine and helps make sense of the complexities of ubiquitous location-based data, NGA will provide geospatial intelligence that only our expertise can deliver, and that our partners can incorporate with their own geospatial data and intelligence workflows.



National Geospatial-Intelligence Agency
7500 GEOINT Drive • Springfield, Virginia 22150

nga.mil

natlgeointagency

@nga_geoint

Approved for public release 14-472.

40573
August 2014