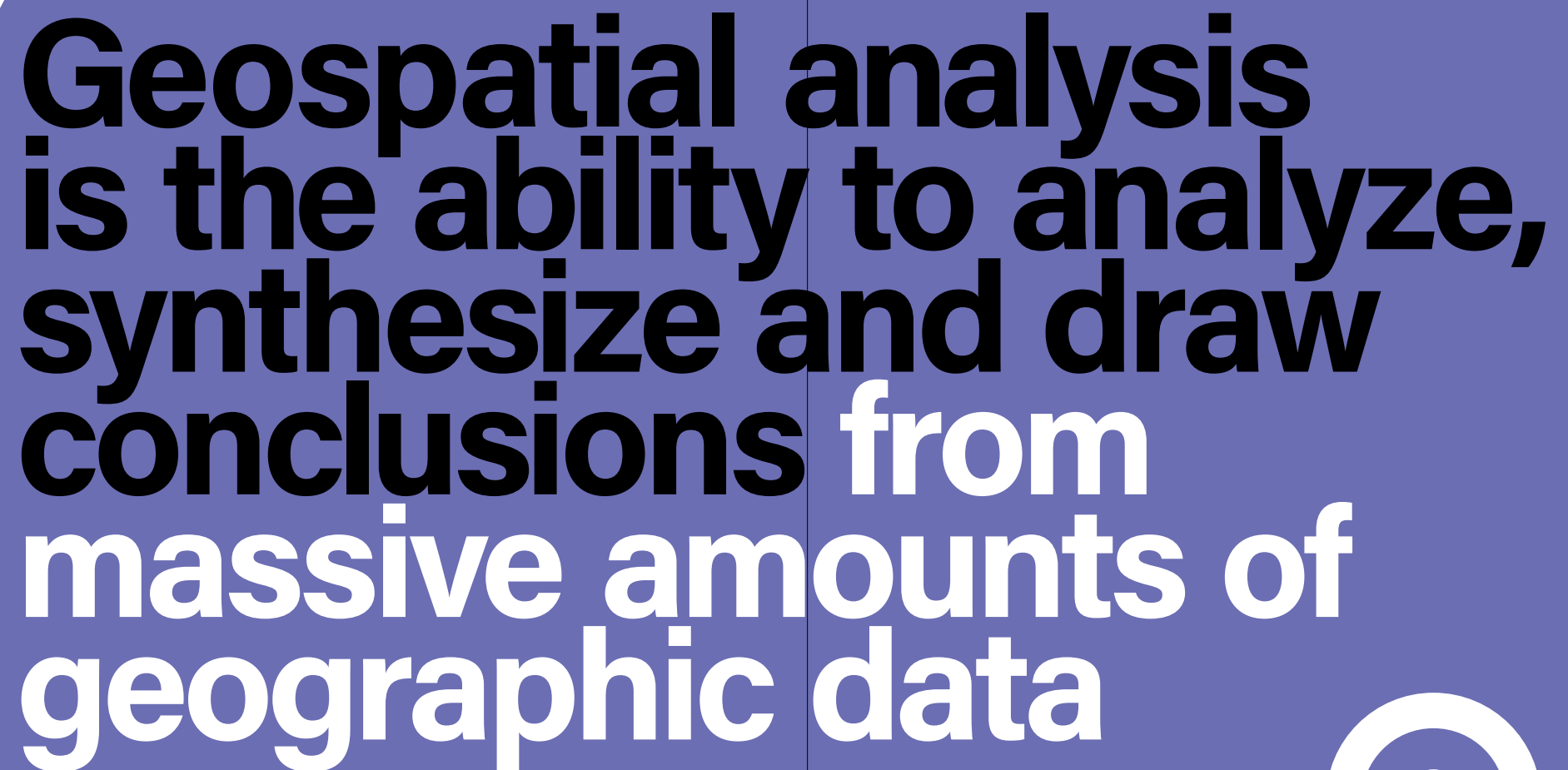GovLoop Pocket Guide
**2016**

# Geospatial Analysis: Breaking Down What You Need to Know

# Geospatial analysis is the ability to analyze, synthesize and draw conclusions from massive amounts of geographic data

# Foreword from MarkLogic

> Analysts have been re-wired to think beyond static maps and are now able to integrate large variety of data sources in real-time. They can also begin to capture complex relationships between concrete and abstract concepts, and objects to improve mission outcomes.
>
> **—Dr. Idriss Mekrez**

From the 15th century Age of Exploration maps, to today's interactive mobile navigation maps, geography has been a key driver for policy decisions or geopolitical events. With the emergence of geographic information systems (GIS), maps have now become electronic documents exceling at describing Earth's physical attributes, terrains and features. Vast geographical domains such as Human Geography (the study of communities, cultures, economies and interaction with the environment) are mostly captured textual descriptions, articles and spreadsheets. There are countless numbers of data sources today – blogs, drone video feeds, mobile photos, and more – the disconnect between well-organized GIS static maps and the reality of unruly, real-time information from new data sources continues to expand.

How can all aspects of reality be captured in a unified view to allow analysis of the world as it is, at this moment in time to make better decisions, faster? As government agencies integrate information from various sources such as news media outlets and crowdsourcing – how can they determine who to trust? How can they merge geospatial information? How do they disseminate it securely?

If data integration challenges were not daunting enough, hyper connectivity within a region and globalization are requiring the capturing and understanding of very complex relationships between people, events, and locations. Everyday events reinforce the importance of relationships while trying to understand world events. For example, the United Nation used historical maps to predict refugees' migration during conflicts or natural disasters. After major events, refugees used well-established corridors, major roads and navigation. The United Nations could predict where to build camps to assist and process refugees weeks before the first population movement. But today, maps and static geospatial analysis methodology is not the same. As illustrated with the Syrian's refugees crisis, groups of well-connected and well-informed refugees use their mobile phones to receive the latest information about checkpoints and closed border crossings. Their behaviors changed in real-time as they explored new routes to reach their destination. This meant some camps were left empty, and others were at overcapacity. This dramatic example contrasts how a simple geospatial analysis problem becomes insurmountable due to the inability to integrate the impact of news feeds shared across a community.

The rise of innovation in government is enabling agencies to implement cutting-edge, state of the art technology solutions to address the harsh challenges faced around Geospatial Analysis. Analysts have been re-wired to think beyond static maps and are now able to integrate large variety of data sources in real-time. They can also begin to capture complex relationships between concrete and abstract concepts, and objects to improve mission outcomes.

Throughout this pocket guide, we will discuss the main topics of the GIS revolution, review practical case studies, and leave you with the conviction that we are ready to make a change in the way we store and access geospatial data to capture and analyze the entire world as it is – in all it richness and complexity.

# MarkLogic®

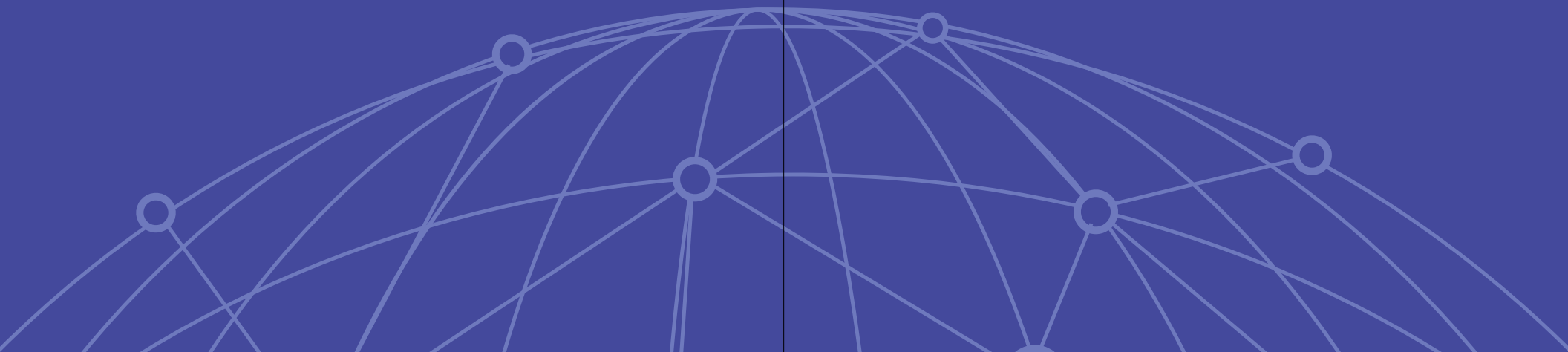# THE WORLD IN CONTEXT

Securely Manage, Discover, Share & Enrich Data in Real Time

The World's Best Database for Integrating Data From Silos

www.marklogic.com

# Contents

# Executive Summary

Since the introduction of the Geographic Information System (GIS) in 1968, its uses have evolved and expanded – from rendering highly accurate maps to identifying the shortest route to your next meeting – and its audiences and value have grown, too.

With the explosion of geospatial data emerging from the web, mobile applications and the proliferation of sensors and drones, however, the next era of GIS innovation has been held back by its underlying data management technology. Simply put, geospatial analysis cannot keep up with rapidly changing data environments.

In recent years, agencies have been creating and collecting massive amounts of unstructured information – and efficiently aggregating and integrating all data into GIS has been a challenge. A high-fidelity representation of the real world can seem unattainable – historical context, human activities and abstract concepts are often absent from current GIS views.

Failure to accurately capture the world and its events, people, places and other information in the proper context can lead to huge risks. In addition to decreased productivity and higher costs by system redundancy and manual operations, it can also negatively affect decision-making. Agencies need to filter through the noise and turn volumes of information into actionable insights.

The new generation of GIS needs to quickly and efficiently capture and manage a multitude of data types, including unstructured data such as e-mail, text messages, photos, video and social media. The ability to analyze, synthesize and draw conclusions from the massive amounts of geographic data collected and created today can help create a smarter, more efficient public sector that delivers better and faster solutions to its citizens.

This quick resource guide will help you better understand this new type of geospatial analysis. We'll discuss an overview of geospatial analysis in the public sector, why it matters, how we can be better at it, as well as case studies that will help you capture and analyze the world as it is – in all its richness, complexity and velocity.

So let's get started. First up: an overview of geospatial analysis in government.

# Overview: Geospatial Analysis in Government

This section will apprise the reader of the definition of GIS and why analysis of it is so important to government; explain some current examples and brief uses cases; and help set the context for why geospatial analysis and capabilities in government matters today more than ever.

## What is geospatial analysis?

It's no surprise by now: Big data is changing the way government operates. With the huge amounts of data that agencies are collecting, managing and storing – they are devising strategies and approaches to utilize and create services geared to the public and internally within the agencies that are using this data.

But there's something often missing from this huge amount of data collected and created by government agencies: proper context and meaning. Having huge amounts of data means nothing if you can't properly understand it, know from where it came from, draw conclusions from it and make good decisions based off of it.

That's where geospatial analysis comes in.

To understand geospatial analysis, first you have to understand spatial data. Spatial data is geographic – and is represented as a "feature." A feature is any static object such as a river, building or street on a map. This geospatial data contains information about the locations and shapes of geographic features and the relationships between them.

Analysis is the act of understanding that data and deriving information and conclusions from it. Simply put, geospatial analysis is how we understand our world by location — mapping where things are, how they relate, what it all means and what actions to take.

Take spatial analysis of crime data as an example. If you can assign a location point or polygon and geographic information to a series of robberies, spatial analysis could help you understand the underlying patterns of how, why and when those robberies might be happening, and help guide you forward in how to deal with them.

# Why does geospatial analysis matter to government agencies?

Currently, most data generated by government agencies has geospatial components. Everything from health records to crime data to transportation routes to population growth can be (and often is) plotted on a map. Today's geospatial tools would be the envy of many an ancient civilization that struggled to understand the world's geography.

The earliest example of geospatial data having a critical public outcome is the story of Dr. John Snow in London in 1854. At that time, a terrible cholera outbreak was decimating the city. Authorities, who believed cholera to be airborne, struggled to understand the source and how to stop it. Dr. John Snow believed that the disease was waterborne – he had only to prove it. To do this, he mapped each cholera death. As the dots clustered, they eventually revealed a centrally located water pump that he believed to be the source of the outbreak. He was able to convince the government of this fact

– and he was correct. The water pump was shut down, the sewer system re-engineered and further outbreaks were prevented.

That was in the 19th century. Today, well into the 21st century, having even more information at our fingertips, and a massive amount of theoretically life-changing data, government agencies should be fundamentally changing due to the sheer potential of geospatial analysis. This is why geospatial analysis matters today for the public sector – it allows the government to think about tying seemingly disparate data points together by location, to reveal patterns and insights that can transform information delivery and citizen engagement.

This change is not happening, or it's not happening as effectively and efficiently as it should. Why? Because there are a number of obstacles to well-executed and life-changing geospatial analysis that government agencies still struggling to overcome.

# Common
## challenges

More data is being collected and created than ever before – but that's not the problem. Every time a person sends a tweet, takes a picture, sells a home, ships a package and more, it generates data. Most of this data, also has a geospatial or location aspect to it. Massive amounts of geospatial data are available to us and ripe for the picking. But as it turns out, it's all about the database and its ability to properly manage massive amounts of data.

A major problem that the public sector faces in executing effective geospatial analysis is that most existing geospatial databases are unfit to manage unstructured information: documents, observations, human geography or situational awareness. Too often a geospatial database plus a spreadsheet and a text editor becomes the de facto system of maintaining real-time data, leaving analysts to manually correlate and search geospatial features and institutional knowledge in separate locations.

With most geospatial teams managing hundreds to thousands of data sources, the problem is compounded by having to search a geospatial system and numerous feature-related flat files. It is impossible to exploit all of an organization's geospatial data in a unified perspective with current geospatial systems. You might say analysts are wasting too much time trying to find the haystack, rather than trying to find the needles within the haystack.

Other challenges to executing on the potential of geospatial analysis include, but are not limited to:

## 1

From news stories, reports and organizational messaging, today's information sources exist in a variety of unstructured formats. Only after complex transformations and batch processing (often caused by constraints associated with relational databases) can an analyst exploit these sources with a geospatial tool

## 2

Current systems only offer the ability to perform geospatial and temporal searches on highly structured geospatial information

## 3

A lack of skilled workers who know how to perform geospatial analysis

## 4

The need to maintain the static geospatial information in GIS systems and all the dynamic, unruly data in Excel or Word documents

## 5

Inability to capture a fast-changing reality or exploit all the information in a single point of access, and loss of pedigree, provenance and historical contexts

## 6

Inability for analysts to comment, rate and edit GIS features in a non-destructive manner

## 7

Challenges around granularity of the GIS data model. For example: Security is limited by layers, so what if a full layer is public information and only one feature's parameter needs to be secured at a higher level?

The solution is a high-fidelity data integration approach capable of modeling complex, real-time geographic information domains such as human geography, structured observation management, damage assessment and disaster recovery. The solution must place geography in context by answering: Who, what, when and, perhaps most importantly, where.

# Industry Spotlight

An interview with Dr. Idriss Mekrez,
Chief Technology Officer, Public Sector, MarkLogic

"We are helping the public sector better understand and capture the world in context to improve government services and national security."

**—Dr. Idriss Mekrez**

## Understanding how to capture the world in context

When trying to answer critical mission-need questions or build new applications, organizations are faced with the challenge of making sense of all of their data. They likely have information distributed across disparate data silos that tend to be rigid and tailored for specific uses. And, even within data silos, information is often disjointed and devoid of meaning.

Nowhere is this more apparent than in the public sector. Most government agencies have been using geospatial tools and analysis for decades. But what they've been doing with those tools is capturing static objects and static location – like developing maps to capture roads, building footprints, location of citizens and more.

Today, most government agencies have to deal with new data sources that are coming to them almost daily – social media, open data, sensor data, military intelligence or data from a sister agency. And the biggest challenge? Currently, the constant stream of data the government is receiving doesn't fit in current geospatial

tools. Geospatial tools were designed to mostly deal with static and tangible map features – not to provide situational awareness of the latest information, across all the real-time data sources they are receiving.

"The challenge the public sector faces is they need to constantly update their maps and capture all their unstructured information and documents in a variety of systems, and then they have to manually correlate the information by hand," said Dr. Idriss Mekrez, Chief Technology Officer, Public Sector, MarkLogic. "So this is where having a 360 degree view of all your data comes to play."

The public sector needs a "360" degree view all data – persons, places, events or things that are important to its mission – everything from citizens, patients, finances, battlefield resources, oil wells – the list is endless. In order to obtain a complete view of all data, you need a platform that can integrate all data from various silos. That's where the MarkLogic Geo360 solution, which offers a complete view of your geospatial data in context, comes in.

"With the **Geo360 solution**, a government agency will be able to capture and secure geospatial information coming from multiple sources, and show the information in all its richness, within a single geospatial view,"

Mekrez said. "So people are able to not only look at the location but also directly access documents, and have all the historical context that is associated with that information."

Here's how it works. The **Geo360 solution** is a high-fidelity data-integration approach capable of modeling complex, real-time geographic information domains such as human geography, structured observation management, damage assessment and disaster recovery. The solution uses a flexible data model approach to store, enhance, analyze and disseminate dynamic geospatial features and things of life (i.e., "objects"). It places geography in context by answering who, what when and where.

With the **Geo360 solution**, a geospatial feature is combined with an object or entity – which means users can have multiple views of the geospatial data. For example, a "feature" such as the U.S. Capitol can be viewed as three objects: a building, a tourist attraction and the seat of the U.S. Congress. Additionally, a geographic location can be viewed as multiple points, lines or polygons. In current relational GIS databases, multiple locations need to be stored separately as features in the database, while objects, attributes, values and metadata are stored in multiple files in one or more different systems, buried in numerous documents. But with MarkLogic, all the information is stored in one database.

Mekrez gave an example: "Say you are a first responder, and there is a fire or an earthquake affecting a large building. It may

have collapsed, or caught fire. When you arrive on location, and look at your typical mapping software that government agencies are using, you may only have the address, and maybe the footprint of the building."

When looking at the same location using the **Geo360 solution**, however, first responders can see more than just the address and footprint of a building – a list of tenants, blueprints and floor plans of the building can be attached to a location. First responders could quickly come to the conclusion that within the building they have 20 tenants, and they could see on which level each tenant is located, and the quickest route to get to each person safely. Having all information within one screen allows a better response to the emergency.

"And if there is conflicting information, all information can be displayed to them, securely with all the sourcing information, allowing them to make better decisions," Mekrez said.

Making quicker and well-informed decisions with a full view of all of the information you collect is what MarkLogic is all about. Their solutions deliver an accurate representation of the real world, with increased trust and security, allowing analysts to directly capture and relate their geographic knowledge on one platform.

"We are helping the public sector better understand and capture the world in context to improve government services and national security," Mekrez said. "And it all starts with better geospatial intelligence."

# Learning From Others:
## Geospatial Analysis Case Studies

Now that we've covered the context for and challenges to geospatial analysis, let's look at examples for inspiration and to help you build a case for it at your agency.



### Connecting Fairfax County police and citizens

Fairfax County, Va., is one of the largest counties in the nation, with more than 1 million residents. The Fairfax County Department of Information Technology (DIT) uses modern information technologies to improve citizen access to government information and services, including a wide variety of mapping data from the county's geographic information system (GIS). The goal of the DIT is to provide the community convenient access to appropriate information and services through technology.

Fairfax County wanted to upgrade its existing Police Events application to allow the public to quickly search, filter and map police service calls by type, location and date and access preliminary police investigation data from Fairfax County's 9-1-1 call center. For example, a new family moving to the area might want to view which school district has the least amount of sex- or drug-related offenses. With more than a half-million data points from all over the county, searches were overwhelming the existing relational database. To search GIS information, Fairfax County DIT had to overcome many challenges – having

multiple data sources with unique schemas; data complexity and unstructured data; slow and difficult searches due to the amount of data and an inability to easily add new datasets as they were identified.

Fairfax County chose a solution built atop the MarkLogic database platform because it gave DIT the ability to make more complicated queries to its GIS database while allowing both county employees and residents quick and easy access. The application improved query times and search speed, greatly improving user experience on location-based searches, as well as making more public data accessible to citizens. The new system returns searches in sub-seconds, whereas the outdated relational database had to search through data from multiple tools for each search, causing simple searches to take 10 seconds on average.

Today, county residents are using the Police Events application to search through the 400,000 criminal events spanning over past decades. The search can be constrained by date range, type of event and several different geographic boundaries (i.e., type of location and location, distance and more).

Fore more information, please visit **www.marklogic.com/customers/fairfax-county-virginia/**

## Faster, more flexible object-based intelligence

Meeting government-wide mandates for transforming agency systems and processes to become more agile and flexible is particularly challenging for defense and intelligence agencies. With so many teams of specialists studying different aspects of evolving situations, these agencies' data, discovery and information fusion is, by nature, highly non-linear.

Agencies are moving toward cross-agency analytic modernization, increased collaboration and Object-Based Intelligence (OBI)/Object-Based Production (OBP). OBP is an approach that aims to increase information integration and create a common landing zone for data that crosses organizational and functional boundaries.

At a federal geospatial agency, the ability to quickly produce complete geospatial analysis reports is crucial to day-to-day operations. The mission includes answering targeted, location-based inquiries that rely on terrain, demographic and other data. Analysis reports are then shared across the enterprise. The agency identified the need to increase

operational agility and flexibility, as well as the introduction of cutting-edge technologies into tradecraft, as key objectives in a recent strategy document.

Unfortunately, teams at the agency traditionally stored their data locally on shared drives or individual analysts' "shoe boxes." This process forced analysts to guess who had the data they needed for a given project, make an inquiry and request access before they could even start their analysis. The poorly integrated information systems, tools and data formats reduced the already limited time available for deep analysis and lengthened response times for important inquiries.

The agency implemented MarkLogic and Berico's Rivium smart data discovery pilot (an intuitive web-based data visualization, exploration and collaboration platform) to fully enable OBI and streamline mission-critical content. MarkLogic's data management features, flexible data structure and enterprise security allowed the agency to consolidate all kinds of unstructured data into one searchable database while reliably enforcing

data policies and security settings. Analysts use this platform to easily access and analyze the data. It not only provides powerful faceted search across the entire enterprise and external systems, but also the ability to quickly visualize and filter data. Data views include geospatial, heat mapping, social network graphs, charts, tag clouds and traditional tables.

Rivium's collaborative workspaces encourage better sharing across teams. Multiple analysts can work on shared projects using the same space and referencing the same data sources without creating duplicate, "shoe-boxed" data. As each analyst adds notes, comments and connections in the course of their own analysis, everyone in the agency benefits from the curated knowledge.

MarkLogic and Rivium have helped agency users maximize their time on data-driven assessments and analyses to deliver faster, better insights for decision-makers.

Fore more information about Rivium, please visit: **rivium.io**
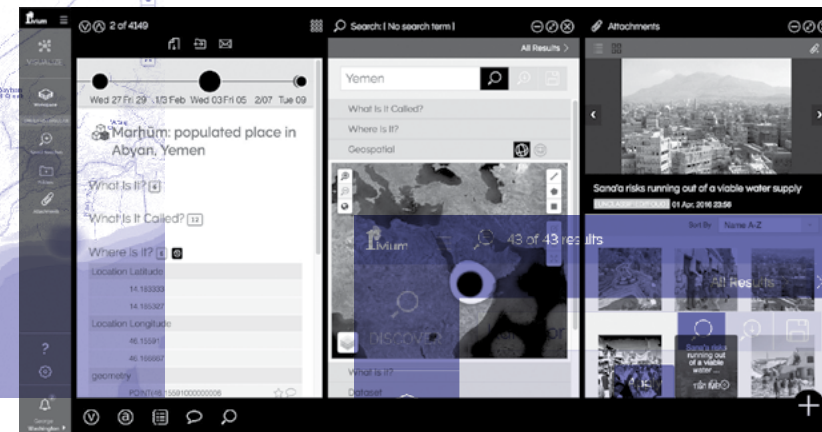
## Combatant command speeds response to rapidly evolving data

Understanding the rapidly changing dynamics of combat situations is one of the toughest intelligence challenges. A Combatant Command continually monitors all kinds of networks – individuals, groups, governments and organizations. Some networks are homogenous, but many are not. Some are tightly located in close-knit communities, while others are dispersed in far-flung, transnational syndicates. Information about these networks can include strategic-level expertise, regardless of geographic location or tactically oriented data about a location or subset of a network.

For the Combatant Command analysts, the mission and networks are evolving constantly, often outpacing the time it takes to build network diagrams and search multiple data sources to discover new links and patterns. The slow process of manually building network maps, disambiguating names and entities and searching disparate data sources to discover links and patterns hampered analysts' ability to respond quickly to developing situations.

Like the geospatial agency, the Combatant Command needed to expedite its current processes to deliver answers faster. It turned to MarkLogic and Rivium for a better analysis solution. MarkLogic's Enterprise NoSQL database helped the agency consolidate access to its diverse datasets. The new system made it faster and easier for analysts to search the data, identify network entities and stay up to date on the latest knowledge.

Now teams can query all their databases of interest and return results in any data format in a customizable view that includes the identities of and links to other teams that have made similar searches. Object-based intelligence is easier since tags link people and places to other essential information, such as financial records, videos and pictures that auto-populate in the structured summary product. The platform provides a near-real-time picture of complex links between people, places and things. The agency makes better use of its data, and teams can collaborate more directly to save time and deliver better results.
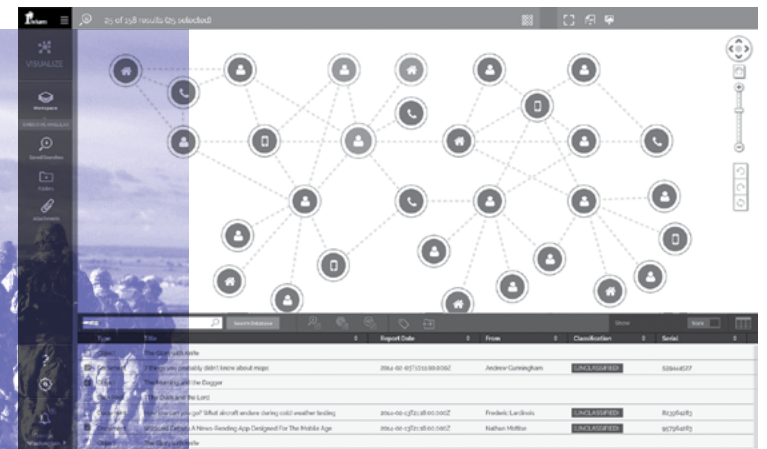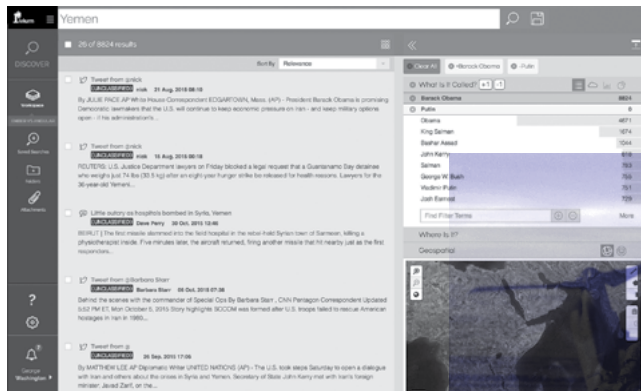
# Geospatial Analysis Cheat Sheet

This takeaway section will highlight the main challenges to geospatial analysis, tips to get the most out of your data and questions to ask to get started.

# Main challenges

You can't overcome challenges to your use of geospatial intelligence unless you first know what they are. Here are the main obstacles you can expect to face in your geospatial intelligence journey.

## 1

### Current geospatial tools don't fit current government needs

Most of the current tools the public sector has at hand have been designed to mostly deal with static features, not to provide a situational awareness picture of all the latest real-time information. This means the public sector must constantly update its maps and capture all its unstructured information and documents in a variety of systems, and then manually correlate the information by hand.

## 2

### The complexity of temporal information

Most maps are very static. For example, if you look for your agency headquarters on Google Maps, it will give you the latest address. But let's say you moved three years ago. You have completely lost the historical content and the fact your agency was ever at a previous address. So if you are looking for another document that referenced your agency address, you will not know how to link that information, because the mapping software is only providing the latest information. This means all historical context is lost.

## 3

### Loss of source of information sources

One of the main challenges for government agencies is that when they bring geospatial information into their mapping or GIS software, they lose the source of the information — there's no way to identify where it came from. For example, during the Ebola outbreak, a lot of agencies were getting data from different NGOs that was very reliable, and they could display it in a map. But a lot of local news outlets in Africa pushed inaccurate information, and some of that information made it into those GIS systems. New mapping software must let you track the source of the information and manage trustworthiness.

## 4

### Different security levels

Not being able to secure information properly is preventing government agencies from displaying all the information in a single view, because all the data sources may not be at the same level of classification. This result is poor data sharing as agencies default to not share information if a fraction of it is confidential.

# Tips for better geospatial intelligence

Get started with these quick pieces of advice from Dr. Idriss Mekrez, MarkLogic's Public Sector CTO.

## #1

**"Start to update the information management system to be able to display structured and unstructured information within your geospatial tool,"** Mekrez advised. **"And in order to avoid adding information that doesn't match the reality of the real world, update your systems to display all current and historic information, securely within a single application. This will greatly improve the ability to make the right decision at the right time."**

## #2

Consider search as a first-class citizen within your GIS application. **"In most mapping or GIS applications, people tend to navigate by moving the cursor or zooming in and out of a map,"** Mekrez said. **"But when you start to bring in documents, media, blogs or crowd-sourced information within your mapping and GIS application, you need to filter information using advanced search capabilities."**

## #3

Look seriously at security. **"In most cases, government agencies will operate at multiple security levels, and it makes it very hard for analysts to have access to all the information in one place, because they have to log into multiple systems to have a full view of what's going on,"** Mekrez said. **"If you can properlyaccess different security levels within one geospatial platform, that greatly simplifies access to information."**

# Questions to ask

Three questions your agency should be asking to put yourself on the right geospatial intelligence path.

## 1

Do you have the latest information in all your systems to make the best decision or do you experience lag times when new data needs to be loaded?

## 2

Are your analysts and decision-makers having to use multiple systems and correlate information manually in order to gain situational awareness or know what's happening in a particular location?

## 3

Are you able to bring the latest sources of information into your current mapping or GIS system without revisiting and revising your database schemas?

# Thanks to MarkLogic for their support of this valuable resource for government professionals.

**MarkLogic**

**BERICO**
TECHNOLOGIES

**govloop**

**About MarkLogic**

For over a decade, organizations around the world have come to rely on MarkLogic to power their innovative information applications. As the world's experts at integrating data from silos, MarkLogic's operational and transactional enterprise NoSQL database platform empowers our customers to build next generation applications on a unified, 360-degree view of their data. Headquartered in Silicon Valley, MarkLogic has offices throughout the U.S., Europe, Asia, and Australia.

For more information, please visit
**www.marklogic.com**

**About Berico Tehnologies**

Berico's interdisciplinary team of engineers and analysts develops user-centric and transformational data integration and analytic solutions. Our scalable end-to-end solutions and elegant user interfaces harness complex data and simplify data discovery.

For more information, please visit
**bericotechnologies.com**

**About GovLoop**

GovLoop's mission is to inspire public sector professionals by serving as the knowledge network for government. GovLoop connects more than 250,000 members, fostering cross-government collaboration, solving common problems and advancing government careers. GovLoop is headquartered in Washington, D.C., with a team of dedicated professionals who share a commitment to the public sector.

For more information about this report, please reach out to **info@govloop.com**

## The future of
## geospatial analysis

Despite myriad challenges, geospatial analysis is gaining strength every day, and the lessons learned by its pioneers are now coming to the forefront. It's important to understand what geospatial analysis and its possibilities could mean for you, your agency and the citizens you serve — and how it can improve your mission.
We hope that this guide helps you better comprehend the issues at hand, and through advice from the best in the field provided you with proven tools to reach success with your initiatives.

## govloop

1152 15th St. NW Suite 800
Washington, DC 20005

M (202) 407-7421
F  (202) 407-7501
www.govloop.com
@GovLoop