



AI-POWERED ANALYTICS FOR THE COMMON GOOD: Strategies for Health and Human Services Agencies

INTRODUCTION

The coronavirus pandemic has highlighted the need for accurate, real-time data and analytics for public health agencies not only to track the spread of the virus, but also to limit its impact through proactive case management, and to make informed decisions about stay-at-home orders and other restrictions. It also has shed light on longstanding challenges with legacy systems and processes that keep governments from bringing together the wide range of data needed to make critical decisions — including, sometimes, life-or-death decisions — on behalf of their citizens.

“Covid-19 has established how poor our data infrastructure is,” says Dr. Suman De, principal consultant and head of government healthcare analytics solutions for Infosys Public Services.

“We are missing the boat on understanding the data science of health and human services programs,” says Otto Doll, senior fellow at the Center for Digital Government (CDG) and former CIO in Minneapolis and South Dakota. “As senior leaders, our people are telling us that we don’t really understand the most useful information. I think that’s a big message that we’ve got a lot of work to do.”

Data analytics are vitally important, but the complexity of data management means states need solutions that utilize artificial intelligence. AI-powered analytics can fast-track data management and analysis across multiple data sources at scale. That allows health and human services leaders to discover insights and make better use of their data for proactive decision-making to improve public health outcomes.

Data analytics are vitally important, but the complexity of data management means states need solutions that utilize artificial intelligence. AI-powered analytics can fast-track data management and analysis across multiple data sources at scale.

But the virus is merely the latest example of how health and human services agencies are finding it difficult to make full use of the vast troves of data they collect. “Agencies realize that despite holding a lot of information, they are not able to proactively act on it,” says De. “Proficiency in analytics is key to improve outcomes, but it is also gated by the barriers of affordability, technology and skillsets.”

Armed with the right kind of data, governments can uncover hidden insights about their populations, helping deliver the right services to the right people at the right time. Advances in data science technology, including machine learning and AI, are making it possible to make sense of rapidly growing amounts of data in new ways, more quickly and at a lower cost. But too often, governments remain behind the curve.

WHY HHS AGENCIES ARE STUCK CRUNCHING THE NUMBERS


From the earliest days of data collection and analysis, governments have tried to use data to understand and explain what’s happening in and around them.

“Too often, we were trying to do analytics with ledger sheets, paper and pencil,” says Phil Bertolini, co-director of the Center for Digital Government (CDG) and former deputy county executive and CIO for Oakland County, Mich. Agencies have improved somewhat since those days, but they’ve still got a long way to go.


Today, ledger sheets have been replaced by big and ever-growing stores of data. That’s especially true in the areas of health and human services, which collect information

THE ANALYTICS CONTINUUM

 **Descriptive analytics:**
Here’s what is happening

 **Diagnostic analytics:**
Here’s why this is happening

 **Predictive analytics:**
Here’s what will happen

 **Prescriptive analytics:**
Here’s what to do to achieve a desired outcome

 **Cognitive analytics:**
What do we now know?



including vital records, hospital admissions, disease incidence and characteristics, drug prescriptions, immunization information, patient treatment records, claims information, health survey records, child and family services information, disability, and behavioral and mental health data. There are new sources as well, such as data from wearables and other connected devices.

But significant challenges prevent health leaders from effectively aggregating and analyzing that data. For one, all those pieces of information are typically scattered across a wide range of databases and agencies.

This segmentation of data will never give leaders a comprehensive understanding of what’s going on.

“Any kind of analytics and reporting based on silos will give you a very narrow view of things,” says Anthony Farinha, senior director for business development of Couchbase, a cloud database technology provider. Doll is even more direct: “If the public understood what kind of decisions are being

made by a department or agency relying solely on the information they collect themselves," he says, "they would be shocked."

Some governments are trying to share data across agencies, but they're often stymied by concerns over privacy, security, siloed data, data stewardship and interoperability. Other systemic challenges, including bad data quality, poor data literacy and inadequate analytics maturity, hamper agencies' ability to use data to drive decision-making.

As a result, many organizations are stuck at the earliest stages of the analytics continuum: using data merely to describe what's happening in the present, rather than projecting future trends or developing appropriate responses (see box on previous page). But officials are eager for more sophisticated analytics, says Doll. "In government particularly, senior leaders are champing at the bit for the more prescriptive types of analytics. We know where we are and where the trend lines are, but how can we get to much better outcomes?"

The coronavirus outbreak represents a highly visible need for health agencies to move beyond the descriptive analytics currently in use — how many cases, where they are, recovery and mortality rates — and leverage predictive analytics. More advanced use of analytics would help government leaders gain critical insights to proactively restrict the spread of the virus by predicting the transmission rate, the epicenter and the most at-risk populations, while limiting the impact on citizens and the economy.

"If we're able to coordinate an understanding of how the virus is spreading over time with transportation, population density and social determinants of health, then I can have a better way to claw back and target the lockdown to take care of the hotspots while keeping other areas open," De says.

Beyond that, multiagency data can identify which populations need more assistance such as food and unemployment benefits. In addition to the pandemic, there are countless

BIG DATA BY THE NUMBERS

The explosion of new data, most of which is unstructured, makes it even more important for health agencies to engage AI-powered solutions.



163 zettabytes:

Amount of data in the world by 2025

80 percent:

Amount of that data that will be unstructured



55-65 percent:

Annual growth in the amount of unstructured data



1 percent:

Amount of unstructured data that is analyzed or used at all

80 percent:

Amount of analysts' time spent identifying and preparing data for use



Sources: Datamation, IDC, Harvard Business Review



UNDERSTANDING THE DATA PROBLEM

Health agencies face many challenges in trying to improve data use and analytics. Here are some common issues:

⚙️ Data jam:

HHS agencies have complex governance processes and lack the right infrastructure to enable data collaboration across departments and agencies. Information stays in silos, making holistic decision-making difficult.

⚙️ Bad data:

Data quality is a major problem. Pieces of information are often scattered across siloed systems in a range of formats, making them difficult to aggregate, standardize and share. Citizen records are often duplicated. In some cases, relying on legacy technologies raises privacy concerns such as personally identifiable information (PII) exposure, data loss and leakage.

⚙️ Unstructured data:

Much of the data in government is unstructured; it lacks the consistent formatting typically needed to yield meaningful insights. Overall, about 80 percent of new data in the world is unstructured; within government, that figure is still as high as 31 percent.²

⚙️ Data literacy:

Challenges in attracting and retaining employees with high-end data science skills make it difficult for agencies to take advantage of advanced data science techniques and run analytics projects at scale and in a democratized way.

⚙️ Inadequate analytics maturity:

Most health agencies are still in the infancy of their analytics maturity. Analysis is still primarily spreadsheet-driven and doesn't have self-service capabilities. Most agencies that do use advanced techniques are still only able to analyze a fraction of their data.

other examples of how AI-powered analytics could help agencies connect the dots when it comes to offering critical services to citizens. With advanced data science platforms, HHS teams could easily bring together federal Census data, state information on unemployment and SNAP benefits, and local-level public health data. Fusing those and other data sources allows an agency to identify and assist individuals early, before they require more costly interventions down the road. For instance, a pregnant, jobless mother who lives in an urban food desert could be targeted with food assistance and information about healthy eating.

“The biggest thing is personalization,” says Jignesh Desai, a solutions architect at Amazon Web Services. “Running the right artificial intelligence models with the right data could help all public use cases, from managing population health to fighting substance abuse and ensuring payment integrity.”

Without AI-powered analytics, agencies are unable to make those kinds of vital connections. One estimate suggests that analysts spend as much as 80 percent of their time preparing data.¹ That compromises government’s ability to actually analyze that data in any meaningful or comprehensive way, or to use it to improve decision-making. It’s a case of “data, data everywhere,” says De, “but not a single drop of strategy.”

FROM NUMBER-CRUNCHING TO BEING VISIONARY: HOW AI CAN TRANSFORM HEALTH SERVICES

AI-powered analytics represent an entirely different way of engaging in analytics. Taking advantage of that potential requires transformation, both technologically and organizationally.

“The reality is that we have this incredible data but massive boundaries,” says Ryan Levy, chief operating officer of Knowi, a unified analytics platform. “One of the fundamental goals is

The traditional rows-and-columns approach to data that began with paper ledgers still exists as spreadsheets in many modern databases. That makes the process of formatting and structuring data from different sources and systems a complicated and time-consuming task.

tearing down these boundaries to make life-changing decisions.”

Government technology leaders have long struggled with technical debt with their analytics initiatives – the cost, time and skills involved in dealing with legacy and custom-designed systems that addressed specific problems or workflows when they were put in place but are difficult to retrofit or scale to meet today’s needs. As they modernize, agencies must look toward an insightful, granular data ecosystem for trustworthy, timely and actionable analytics. Traditional analytics solutions have limitations that make it difficult to process and act on data from disparate sources. They don’t adapt to changing requirements quickly enough; they don’t connect all the stakeholders in the care continuum; and they don’t deliver a unified experience to those stakeholders. Government leaders should look to emerging technology capabilities that have the potential to address these longstanding challenges in new ways. Among them:

Aggregation and data harmonization from diverse sources,

including internal systems, other agencies, and community and commercial partners. The traditional rows-and-columns approach to data that began with paper ledgers still exists as spreadsheets in many modern databases. That makes the process of formatting and structuring data from different sources and systems a complicated and time-consuming task. “In the past, building the models and schemas has been costly, challenging and a barrier for a lot of these digital transformation projects,” says Farinha with Couchbase. Emerging NoSQL databases do away with relational tables, allowing data to

be stored in multiple dynamic structures and acted upon in different ways – and more quickly than in the past.

“It provides the ability to build the models and schemas in a very dynamic way,” Farinha says. “The big ‘aha’ is that we can provide the ability to query data so it can be acted on in or near real-time.”

The sheer volume of data means it is also critical to automate the data life cycle end to end. That means being able to ingest data in various forms without needing extract-transform-load (ETL) tools. It also means implementing advanced data harmonization techniques that automate key data management processes like intake, validation, metadata management and normalization. And it means being able to automatically create a single repository of data to support purpose-driven business intelligence reporting and advanced analytics.

Automation is critical, given today’s large and growing volumes of data. “Without it,” says Bertolini, “you wouldn’t be able to look at the data.”

A unified, advanced AI analytics platform.

To avoid the equally time-consuming process of crafting new analytic models for each question or analysis, it’s important to draw from a platform that has a common toolkit of models. This platform will leverage next-generation information management techniques to intelligently interpret, analyze and act upon data while also continually building knowledge and the learning curve. It will further automate the process of building, developing, testing and maintaining the predictive models. This eventually democratizes data science skillsets within an agency and provides

As with other government technology, the cloud is becoming a critical component of next-generation analytics. A common platform based on cloud technologies can simplify deployment and access, allowing solutions to scale for any dataset, workload and number of users without impacting performance.

a self-service approach which will enable skilled analysts to draw from existing algorithms and generate predictive insights, more quickly and at a lower cost.

For example, an analyst could use standard ranking models to prioritize which data points — age, symptoms, doctor visits — carry the most weight in a predictive model. A unified platform “enables the end users with the knowledge of data variables they need to be aware of to understand the model behavior,” De says. In the current era of data democratization, frontline employees’ data skills have evolved. Unified analytics enable agencies’ business analysts and subject matter experts to easily merge data science and data engineering for rapid predictive model development on their own, without depending on expensive and hard-to-find data scientist skillsets.

Another important aspect will be the platform’s ability to go beyond insights and generate actionable recommendations, which can be shared with healthcare professionals in the care continuum, enabling them to collaborate and deliver the right care to the right person at the right time.

Visualizations and easy access to data.

Visualizations are particularly important for senior decision-makers. Leaders “need the trends and the highlights,” says Bertolini, “not a spreadsheet with 30 columns and 40 headers.”

They’re also particularly important in communicating health information beyond government by presenting information in easy-to-understand ways, to “put the power of data-driven insights in the hands of everyone

who wants to understand what’s happening in their state or their county,” says Knowl’s Levy. “[Tools] to visualize the data provide the ability to take complex data from disparate sources and tell a valuable story in a simple manner.”

Finding ways to communicate the real-world meaning of data is critical, says Harvard economist Raj Chetty, who has used big data to help illustrate the impact of poverty in his research. “If what we’re doing is important for the world, we should make it accessible to the world,” he told an interviewer in 2018.³

Simplifying access to data goes beyond visualizations. It also involves natural language processing (NLP) or search-based analytics to allow users and citizens to ask questions using plain English, without knowing the specific terminology or parameters of a query — and still get accurate answers. This eventually will enable non-technical users to query the platform for answers and quickly create purpose-built dashboards and visualize data in a manner that is easy to interpret and ready to make decisions.

“It really makes it simple for people to ask questions without knowing how to write a query or understanding the underlying architecture of the data,” says Levy. “Ultimately, that allows anyone to see their data how they want, where they want it and when they want it.”

Leverage the cloud for scale, lower cost and seamless data collaboration.

As with other government technology, the cloud is becoming a critical component of next-generation analytics. A common platform based on cloud technologies can simplify deployment

and access, allowing solutions to scale for any dataset, workload and number of users without impacting performance. It also enables agencies to adapt to rapidly growing and changing requirements. Given privacy laws, cloud services also can provide fine-grained control of where data resides and how it is shared. Additionally, it ensures seamless sharing of data with internal business groups and external partners without any need to physically copy, move or deconstruct data. That gives agencies the flexibility to quickly adapt new technology at a lower cost.

STRATEGIES FOR TRANSFORMATION: HARNESSING THE POWER OF AI FOR ANALYTICS

Governments committed to improving their analytics capabilities need to look deeper than new tools. Data-driven strategy should be at the core of any modernization project planning. This will enable them to use data not just as a source of information but as a source of intelligence and insights to drive outcomes with precision.

“You have to have a foundational capability to deal with your data. If you don’t understand the domain — what target variables you have to define, what business problem you have to solve — you don’t understand the context,” De says. “Even with the best data scientists, you won’t get the right outcomes.”

Next, it’s vital to develop a framework to assess where agencies are in terms of their access to data and their capabilities to analyze it.

“You need structure, process and systems to make sure you get the data you need to do really cool stuff,” Bertolini says. “There’s an investment in understanding what you have and what you need, and then getting what you need.”

To move forward, agencies should:

Classify data, which involves asking a wide range of questions about existing and needed data, including:

✔ **What data do we need?**

✔ **Where does the data reside?**

✔ **Is the data clean and of good quality?**

Are records complete, particularly in legacy systems? Historical records may be split between aging databases and paper records.

✔ **Can the data be used for decision-making?**

How will the existing data support a hypothesis? Do I have all the variables present? In particular, does the metadata — the data about the data — include the information that can determine if it's useful in decision-making? Metadata about the time at which data was collected is particularly critical, as some analytics draw from historical data while others, such as critical decisions about a viral outbreak, require it to be up to date.

✔ **What rules govern its use and sharing?**

What consent is required to exchange the data? Given HIPAA and other privacy regulations, it's important to understand how personal data can be used and shared across agencies. For example, personal information that may not be sharable for an individual — such as specific health issues or incidents of opioid abuse — may be usable in aggregate forms if personal information isn't included.

✔ **Are there sufficient cybersecurity and authorization/access controls to ensure those rules are followed?**

Identify new data sources — and ensure those sources understand how their data is being used.

While governments typically have data-sharing agreements with other health organizations, they will likely need to broaden in scope over time. As Stanford researcher and professor of pediatrics Atul Butte told an interviewer a few years ago, “hiding within those mounds of data is knowledge that could change the life of a patient, or change the world.”⁴ Agencies need to “build a culture that

values data sharing,” AWS’s Jignesh Desai says, opening the door for new public-private partnerships and opportunities. “If [governments] can make their data more democratized, a lot of companies will come up with these kinds of solutions.”

Two keys to doing so are clearly articulating appropriate uses for shared data and ensuring confidentiality, he says. Each organization also will require its own data-sharing agreement, which spells out procedures and privacy considerations for data exchange.

Invest wisely in technology. Seek common platforms that span the entire data life cycle — from acquiring and ingesting data to its governance and use in data analytics and storage and archiving. Using multiple incompatible tools, as most agencies have done in the past, creates friction points related to talent gaps; operational inefficiencies; and a lack of collaboration among data scientists, policymakers and other stakeholders. A common platform alleviates that pain.

“From start to finish, the solution has to be purpose-built for each stage of the data life cycle,” Desai says.

AI and machine-learning tools are particularly important. This increasingly sophisticated technology, says De, allows organizations to see “the real art of using data in a transformative way, from records to facts to information to intelligence to insights-driven experience.”

Develop metrics to track outcomes.

For analytics to fulfill its end goal of offering predictive and prescriptive insights, solutions will need to be refined by the results from ongoing interventions to inform future decision-making. Agency leaders will also need a deep understanding of what challenges they are trying to address through data analytics, and what insights will help them achieve faster business transformation by extracting value from the data.

“The technology has to be robust enough to get what I need when I need it,” says Bertolini. “But the ‘why’ is more important than the bits and bytes.”

CONCLUSION

Health and human services continue to grow more complex — and costly — for governments to manage. As is true for healthcare providers, agencies will need to evolve to a more proactive approach of anticipating and responding, not reacting, to the needs of their citizens. That, in turn, will require them to adopt new technologies that allow them to draw insights from a wide range of organizations and data rapidly and efficiently enough to intervene before crises emerge.

Looking ahead, data science and analytics will evolve further, with more niche cognitive capabilities and a far more advanced way of doing things. Organizations will rely on automation for the entire life cycle of their data — from data aggregation and harmonization to predictive model development and insights generation.

Ultimately, AI-powered analytics will transform HHS systems all the way from enterprise strategy discussions to frontline decision-making.

“Putting that power in the hands of healthcare workers makes them feel confident that they have up-to-date, accurate data from which they can make decisions that much faster,” says De. “Automated AI platforms will help government healthcare agencies turn their data into actions and share those actions with healthcare professionals, enabling them deliver proactive interventions that result in the best outcomes.”

This piece was written and produced by the Center for Digital Government Content Studio, with information and input from Infosys Public Services, AWS, Couchbase and Knowi.

Endnotes:

1. https://www.informatica.com/content/dam/informatica-com/en/collateral/executive-brief/big_data_government_ebook_2340.pdf
2. <https://www.techrepublic.com/article/unstructured-data-the-smart-persons-guide/>
3. <https://hbr.org/2017/05/whats-your-data-strategy>
4. <https://hbr.org/2017/05/whats-your-data-strategy>

Produced by:



The Center for Digital Government, a division of e.Republic, is a national research and advisory institute on information technology policies and best practices in state and local government. Through its diverse and dynamic programs and services, the Center provides public and private sector leaders with decision support, knowledge and opportunities to help them effectively incorporate new technologies in the 21st century. www.centerdigitalgov.com.

For:



Infosys Public Services accelerates digital transformation for the public sector. We do this by adapting 40 years of cross-industry experience for the public sector and combining it with next-generation technology solutions, proven delivery models and continuous learning and innovation. www.infosyspublicservices.com



Amazon Web Services (AWS) is the world's most comprehensive and broadly adopted cloud platform, offering over 175 fully featured services from data centers globally. Millions of customer s— including government agencies — are using AWS to lower costs, become more agile, and innovate faster. www.aws.amazon.com



Couchbase provides an enterprise-class, multicloud to edge database that offers the capabilities required for business-critical applications on a highly scalable and available platform. Couchbase is built on open standards, combining the best of NoSQL with the power and familiarity of SQL, to simplify the transition from mainframe and relational databases. www.couchbase.com



Knowi's Unified Analytics Platform has been built from the ground up for modern data architectures. Knowi provides instant connectivity to Any Data (NoSQL, SQL, API), Anywhere and combines AI, NLP and BI to create the intelligence needed for Anyone to make data-driven decisions. www.knowi.com