FROM DATA TO DECISIONS III

Lessons from Early Analytics Programs

NOVEMBER 2013

PARTNERSHIP FOR PUBLIC SERVICE

IBM Center for
The Business of Government

The Partnership for Public Service is a nonpartisan, nonprofit organization that works to revitalize the federal government by inspiring a new generation to serve and by transforming the way government works.

The IBM Center for The Business of Government connects public management research with practice. Since 1998, the Center has helped public sector executives improve the effectiveness of government with practical ideas and original thinking. The Center sponsors independent research by top minds in academe and the non-profit sector, and creates opportunities for dialogue on a broad range of public management topics.

INTRODUCTION

Kenneth Dickie still shudders at the memory. One day in 1979, someone snuck into his secret office in the basement of the [Veterans Affairs Department's] Washington Medical Center. The intruder stacked piles of patients' records around Dickie's DEC minicomputer, doused them with a flammable material and set them on fire. Smoke filled the room, but fortunately for Dr. Dickie and for the future of American health care, an alarm went off in time, and the computer he was using to build the country's first practical electronic medical record system was spared."[1]

It's hard to imagine in today's data-drenched government, but early analytics users sometimes were scorned by agency staff as rogues and renegades. Dickie and a band of other amateur programmers overcame attacks from VA's central computer office by winning over doctors with their fledgling health record software and thereby getting senior leaders' support.

Had Dickie's machine burned or had higher-ups failed to protect the secret programmers, one of today's most widely used electronic medical records, the Veterans Health Information Systems and Technology Architecture (VISTA), might never have been created. Without it, the VA might never have adopted the wide array of data-based performance measures and tools that have helped transform it into a hospital system whose quality and safety scores are among the best in the United States.

Dickie's story is a fitting beginning for the third report in our "From Data to Decisions" series, a collaboration between the Partnership for Public Service and IBM's public sector business analytics and optimization practice. Our focus this time is mature analytics programs, such as the VA's, that started before the terms "big data" and "analytics" were in use, let alone in vogue.

Analytics is the study of data to discover patterns, opportunities and linkages that enable prediction and inform decisions. Data trailblazers like VA have much to teach agencies being pressed to use analytics now. That's why we examined how early programs got started, what sustained them and how data use altered mission-critical programs. We purposely focused on true "mission analytics" programs that apply data-based analysis directly to improve mission delivery or performance.

Government, like industry and commerce, is inundated with data demands, especially with Barack Obama, who has been dubbed "the big data president," at the helm.[2]

On his first day in office in 2009, Obama issued a memo on transparency ordering agencies to put operational information online.[3] In March 2012, he issued a $200 million big data research and development initiative.[4]

And in May 2013, he set a new standard requiring

1 Phillip Longman. *Best Care Anywhere: Why VA Health Care Would Work Better For Everyone*, (Bk Currents, Berrett-Koehler Publishers, San Francisco 2012), Kindle Edition, 23.

2 Nancy Scola, "Obama, the 'Big Data' President," *Washington Post*, 14 June 2013, http://bit.ly/18Ghl3F.

3 White House Office of the Press Secretary, Transparency and Open Government Memorandum, January 21, 2009, http://bit.ly/lfWWOgD.

4 White House Office of Science and Technology Policy, "Obama Administration Unveils 'Big Data' Initiative," March 29, 2012, http://l.usa.gov.1blTsoW.

that government data be online, open and machine-readable, which will make larger amounts available for analysis by the public.[5] His administration demands data-driven program evaluations and evidence-based rulemaking.

Federal agencies, like companies, are susceptible to the deafening hype about how big data will improve productivity and process. But evidence is beginning to show that the return on big data investments to date is less than promised.

For example, a recent survey by Wikibon, an online advisory community of technology and business systems experts, found that enterprises expect a return of $3 to $4 for every $1 invested in big data technology over three to five years, but so far are seeing just 55 cents on the dollar.[6]

While most Fortune 500 companies are deploying or plan to implement big data projects, it appears that many are doing so without tying the masses of data that result to specific business goals. Projects driven by line business departments—not information technology departments—and focused on "small, but strategic use cases," are the most likely to deliver significant value, according to Wikibon.

A September 2013 report on big data adoption by IT consultancy Gartner found that while 64 percent of the organizations surveyed, which included government agencies, had invested or planned to invest in big data technology, 56 percent said they were struggling to wring value from it.[7]

Gartner advised companies to "ensure big data initiatives are tied to organizational goals and processes and demonstrate the insights and value that these initiatives bring to the business."

Unlike today's big data initiatives, federal analytics programs that began years ago had no choice but to provide and demonstrate value. They didn't have the luxury of sophisticated data-collection technology. Gathering and analyzing data was arduous and unfamiliar enough that it was employed only to answer the pressing questions most important to the mission. Programs started out small, discrete and focused on mission-critical results, without today's blazing-fast software and nearly limitless storage technology.

Today managers are tempted to begin analytics programs before determining the mission-essential questions they are seeking data to answer. This is possible because computers and software now can store and analyze data faster and at less cost than ever before.

EXPLORING THE "HOW" OF ANALYTICS SUCCESS

The cases we studied vary by the volume and types of data they collect and the ways they analyze it. Some agencies operate data warehouses and use predictive analytics. Some analyze images of bacterial DNA for epidemiological investigations; others identify insects during import inspections. One relies on human brainpower and experience to synthesize information collected by satellites and processed by physical scientists, statistics produced by African governments and data gathered by social scientists visiting rural markets.

Some of our findings mirror our previous two reports. Successfully changing how mission-critical programs operate always has required sustained leadership attention, for example. And employee buy-in is vital.

What differs in this report is that the programs we studied have been in operation longer, allowing us to offer a deeper look at how they have advanced. They offer models for achieving the support, collaboration, cost-benefit metrics, buy-in and other factors we and others have urged agencies to adopt.

We now experience daily the seemingly magical results of big data: personally tailored recommendations for things to buy; instructions on getting to the websites and stores where those things are sold; medical treatments based on parsing the human genome. Twenty years ago, collecting and analyzing data about Veterans Health Administration (VHA) patients or the conditions likely to cause famine—two cases we examine—was neither easy nor common.

Today's analytics projects often are driven top-down by program managers or agency leaders seeking to comply with administration mandates, husband their resources and take advantage of new technology. Older data-based programs often grew from the discoveries of line employees, who made connections and saw patterns in data after receiving new software or hardware that offered a broader, more organized view of existing information.

The lesson from grassroots-driven older projects is that managers should not overlook the payoff that comes from enabling employees to see and use data organized for their needs. Giving this power to employees inspires insights and a thirst for more data and ways to link it. Those insights can help analytics programs evolve to deliver even more mission improvement.

5 White House Office of the Press Secretary, *Executive Order: Making Open and Machine Readable the New Default for Government Information*, May 9, 2013, http://l.usa.gov/ldSKLUg.

6 Jeff Kelly, "Enterprises Struggling to Derive Maximum Value from Big Data," Wikibon, Sept. 19, 2013, http://bit.ly/1894Yen.

7 Lisa Kart, Nick Huedecker and Frank Buytendijk, "Survey Analysis: Big Data Adoption in 2013 Shows Substance Behind the Hype," Sept., 12, 2013, Gartner Research, G00255160. http://gtnr.it/1adE6P3.

Earlier programs, like today's, needed senior support. One lesson they offer is that turning leaders into allies entails delivering findings quickly and tailoring them to executives' needs, especially when the goal is persuading them to act on analysis that challenges long-held perceptions or suggests that practices or whole programs must change.

Now, as in the past, data can produce unwelcome results. Practitioners confront resistance from those who gather data and stand to be affected by it. That's why it's important to ensure that collectors see the results of analytics and to communicate how using data often can improve the ways jobs are carried out and missions are achieved.

Analytics has always been a collaborative effort. The best programs look to others for additional data sets, funding, ideas and labor. Methods employed by early users for building and sustaining those relationships—from marketing their analytics projects to potential collaborators to using grants to equip and staff partners—offer tested models for today.

And in today's austere budget environment, analytics programs must justify their costs. They can take lessons from older programs' techniques, such as relying on the great power of output and outcome metrics to make the case for mission-critical data projects. The danger is in relying on these metrics exclusively, without also measuring monetary benefits.

When benefits data was hard to collect, some programs turned to academic studies or the work of professional organizations for support. Perhaps the most important lesson is to expend analytical effort not just on mission improvement but also on demonstrating return on investment.

These cases offer lessons in making analytics a default approach for accomplishing mission goals.

Collaborate with other agencies to collect data and share analytics expertise

Save money and effort, and increase the speed of analytics adoption, by acquiring data and services, such as collection, analysis and modeling tools, from other agencies.

Search for existing authorities that allow you to pay for help. They can range from interagency acquisition of data and services under the Economy Act, to provisions specific to your agency, such as USAID's participating agency service and program agreements. Grants are another source of funds for engaging partners. And when analytics programs can help other agencies achieve their missions, consider striking memorandums of understanding so each partner can perform the work that suits it best.

Marketing data-driven products to other agencies also can bring in funds and assistance. Research and development agencies are accustomed to sharing their discoveries to get them into operation, and the approach is worth expanding, both to prevent duplication and to share increasingly tight resources.

Develop data to determine return on investment for analytics programs

Reporting improved outcomes, such as increased numbers of foodborne illness outbreaks detected or enemy combatants identified, is a bottom-line requirement for mission analytics programs. But just reporting better outcomes is not sufficient, especially now that sequestration is compelling programs to compete fiercely for scarce dollars.

Agency leaders need cost-benefit metrics and measures of ROI to prove that data-based efforts compare favorably with other programs during budget reviews. So don't focus on core analysis so single-mindedly that you fail to develop data to demonstrate ROI. Even for mature analytics programs, this has been a struggle.

Lacking cost-benefit data, some programs turned for evidence to academic studies and reports from professional organizations. Other sources can include federal statistical agencies, such as the Census and Labor Statistics bureaus, and industry groups.

Programs that can point to improved outcomes from the use of mission analytics still are challenged in putting a dollar value on those results, particularly when they are measured in terms of costs avoided.

Some program managers are experimenting with methods for estimating the value of cost-avoidance, for example, by using billing records of trusted vendors to estimate costs avoided by cutting off defrauders. Analysts must be careful when calculating ROI for analytics to report not only projected savings, but also projected costs for their data programs.

In some cases, improving ROI estimates requires increased and enhanced data collection. Programs also can employ surveys and audits, and conduct secondary screening, which captures what analytics-based efforts might have missed, to assess the success of analytics programs and compare their results with alternate strategies for achieving the same goals.

Give agency leaders clear, concise analysis and proof of adoption, and results they can use to support data-driven programs

Much as most analytics users wish everyone would immediately understand and appreciate their findings, that doesn't always happen. Presentation is especially important for top officials whose time and attention are limited, but whose support is vital. Data visualization—charts, graphs, maps and models—make analytical findings easier and faster to comprehend.

Agency leaders can more easily absorb key findings that are condensed into tight, sharply written synopses at the top of reports, in PowerPoints and at briefings. That's especially important when the data runs counter to leaders' instincts or requires difficult action or change. Then persistence is as important as presentation, and continual reference to the data can overcome initial skepticism and emotional responses.

Pointing to effective analytics adoption by willing employees can help leaders overcome resistance from groups that feel threatened. To support data programs, senior leaders need to see and understand the results and how they apply to achieving the agency's mission.

Encourage data use and spark insights by enabling employees to easily see, combine and analyze it

Standardize data so users can look across it by time, entity, geography, source and other attributes to find linkages and patterns and share information. Letting intended users test-drive analytics tools and muck around in the data itself enables discoveries that can save time, ease adoption and ensure success.

Watch what users do with data and analytical results and tools. No matter what you provide, they will come up with new ways of using it. That improvisation often points the way to improvements.

If you solicit users' help, be sure to implement some of their ideas and let their insights guide how the program evolves. The best reward is access to data-driven tools that deliver actionable information when, where and how users need it. They don't want to search for data or sign on to multiple systems to gain access to it.

Make sure those who collect data see the results of the analyses so their ardor for collection doesn't fade. Similarly, make sure employees see a direct connection between data analytics and the mission that drew them to the agency.

Leaders and managers should demand and use data and provide employees with targeted on-the-job training

Making analytics standard operating procedure means building it into the agency's culture and climate. It pervades the culture when managers at all levels use data in planning, measuring results, budgeting, hiring and running programs, and when they demand that employees' work activities and requests are data-based as well. That requires on-the-job training in data analysis, calibrated to each unit and each employee's role within it. Sending out data evangelists with analytics expertise to spread the news about data-driven accomplishments and possibilities can entice employees to seek training.

Making data use standard practice also requires special analytics expertise, technology and software. Some agencies have found success by creating analytics centers where data scientists, policy experts and experienced staffers continuously collaborate to develop and refine tools matched to mission requirements. Some are naming chief data officers at headquarters and in bureaus to run their analytics centers and to evangelize data use.

FINDINGS FROM PREVIOUS REPORTS

From Data to Decisions **The Power of Analytics**

The analytics process turns data into meaningful information that program staff and agency leaders can use to make good decisions.

Leadership support and analytics are cornerstones of performance management, which requires supervisors and managers to identify problems, assess progress and share results.

For analytics to become accepted widely, leaders should set expectations and call for accountability.

Nonexperts, whether leaders or line employees, need data that they can access easily, understand and tailor to their needs.

Collaborating with partners and stakeholders enables agencies to share data for analytics use, improving results.

Sharing data requires transparency.

The goal is to foster analytical insights, whether agencies have state-of-the-art data tools or less advanced software.

For analytics to succeed, employees need a supportive environment, training and the encouragement to use and experiment with data.

From Data to Decisions II **Building an Analytics Culture**

To get started with an analytics program, create a team with agency experience, analytical skills and subject-matter expertise.

Craft questions about work processes and other agency activities that will lead to data gathering and improvements by: defining a current process, describing an improved state, focusing on top issues that need to be addressed and agreeing on a desired outcome.

Determine tools or systems needed and show benefits rapidly; then test and refine data requirements.

Communicate accomplishments and next steps clearly and meaningfully to get people on board.

Know and understand the data collected and use it to make decisions.

Encourage collaborative partnerships internally and with other agencies and partners outside the federal government.

Bring in people from various disciplines who will examine data and approaches from different perspectives.

Broaden employees' knowledge and viewpoints and share their program expertise by moving them from program to analytics offices.

Case studies at a glance

U.S. AGENCY FOR INTERNATIONAL DEVELOPMENT

Famine Early Warning Systems Network

The U.S. Agency for International Development funds the Famine Early Warning Systems Network (FEWS NET), which collaborates with the U.S. Geological Service (USGS), National Oceanic and Atmospheric Administration (NOAA), National Aeronautics and Space Administration (NASA), Department of Agriculture (USDA), U.N. Food Security and Nutrition Analysis Unit, World Food Program, other humanitarian assistance organizations and regional governments to provide timely and rigorous early warning and vulnerability information on emerging food security threats in 30 countries.

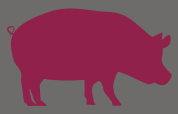
BEGAN: 1986

COST: About $25 million in fiscal 2013

ROI: FEWS NET helps target as much as $1.5 billion per year in USAID Food for Peace assistance to those who need it most.

DATA SIZE: Unavailable

PAGE 8



CENTERS FOR DISEASE CONTROL AND PREVENTION

PulseNet

CDC's PulseNet, a national network of 87 public health laboratories, connects foodborne illness cases to detect outbreaks using a database of more than 500 million isolates of DNA from foodborne bacteria.

BEGAN: 1996

PARTNERS INVOLVED: Association of Public Health Laboratories (APHL), State Public Health Laboratories, Food and Drug Administration, USDA

COST: Less than $10 million

ROI: PulseNet saves an estimated $291 million in medical costs avoided each year.

DATA SIZE: 15 GB

PAGE 10

Observations and lessons

Collaborate with other agencies to collect data and share analytics expertise

PAGE 18

Develop data to demonstrate return on investment

PAGE 21

DEFENSE DEPARTMENT BIOMETRICS

The Automated Biometric Identification System (ABIS), fed by field collection using the Biometrics Automated Toolset (BAT), Handheld Interagency Identity Detection Equipment (HIIDE), Secure Electronic Enrollment Kit (SEEK), and other data

The U.S. armed forces have collected biometric information–ideally 10 fingerprints, iris scans, facial photo and biographic information—from non-U.S. citizens in Iraq and Afghanistan since 2003. Biometrics are used for access to the United States, U.S. facilities and coalition-controlled areas in-country, identification of enemy fighters, forensics (e.g., to identify makers and implanters of improvised explosive devices) and intelligence. Biometrics collected by members of the Army, U.S. Marine Corps, Special Operations Command and international governments are shared with the State and Homeland Security departments.

BEGAN: 2003

COST: $3 billion from 2007 to 2012

ROI: From 2004 to 2012, approximately 3,000 enemy combatants identified, 950 high-value individuals captured or killed, 2,300 detainees denied early release; added 190,000 identities to DOD's biometric-enabled watch list; through February 2011, 538 people seeking asylum in the U.S. turned away due to biometric matches with negative information.

DATA SIZE: 7 million records in ABIS, 4.4 million unique identities

PAGE 12

ANIMAL PLANT HEALTH INSPECTION SERVICE

Agricultural Quarantine Activity System (AQAS)

The USDA Animal Plant Health Inspection Service (APHIS), in cooperation with the Homeland Security Department Customs and Border Patrol, uses AQAS to make risk-based determinations about which shipping containers at U.S. ports to examine for plant-borne pests.

BEGAN: 2007

COST: Unavailable

ROI: Invasive species cause estimated losses of $136 billion annually. Automating the emergency action notification reporting has enabled APHIS to redeploy one full-time analyst into a more valuable role. Replacing duplicative reporting tools with a single analytics solution has reduced costs by 30 percent.

DATA SIZE: Unavailable

PAGE 14

VETERANS HEALTH ADMINISTRATION

Care Assessment Needs (CAN) Score
Patient Care Assessment System (PCAS)

The Veterans Health Administration (VHA) uses the care assessment needs (CAN) score, a weekly analytic predicting the likelihood of hospitalization or death within 90 days and a year, to identify high-risk patients. The patient care assessment system (PCAS) uses the CAN score and a host of other data to enable patient-centered care teams to coordinate services to prevent hospitalizations.

BEGAN: CAN—2011, PCAS in pilot testing since Dec. 2012, rollout slated for 2014

COST: Unavailable

ROI: Patients with CAN scores in the top 10 percent who saw their assigned primary care providers for more than 60 percent of scheduled visits were 10 percent less likely to die or be hospitalized than similar risk patients who did not see their providers during the preceding year.

DATA SIZE: The CAN process collects more than 14 GB of patient-level data (120 unique elements for each score) on 5.25 million primary care patients. CAN and PCAS are fed by the VHA's 80-terabyte corporate data warehouse, which aggregates the electronic health records in the Veterans Health Information Systems and Technology Architecture—approximately 30 million records, 3.2 billion clinical orders, 1.8 billion prescriptions, 2.3 billion vital-sign measurements and 2 billion clinical text notes.

PAGE 16

Give agency leaders clear, concise analysis and proof that analytics are being used to improve mission results

PAGE 24

To encourage data use and spark insights, enable employees to easily see, combine and analyze it

PAGE 27

Leaders and managers should demand and use data, and provide employees with targeted on-the-job training

PAGE 30

Stopping Starvation

Physical and social scientists team up on famine warnings

The world awoke slowly to the 1983–1985 famine in Ethiopia and Sudan. Only after shocking images of masses of starving and dying people were televised did humanitarian aid flow. Even then, Ethiopia's embattled Marxist government was slow to distribute it. Between 400,000 and 1 million people died.

The U.S. Agency for International Development (USAID) created the Famine Early Warning Systems Network (FEWS NET) in 1985 to speed response to future famines and prevent a repeat of Ethiopia's massive loss of life. FEWS NET was created to help USAID deliver its $1 billion in annual food aid where it was most needed and would do the most good, said Gary Eilerts, the network's program manager.

Humanitarian aid is enmeshed in politics. In Africa, where the network focused first, governments can be reluctant to announce a food emergency, which can be read as an admission of policy failure. Aid-giving governments face withering criticism if they drum up assistance for a crisis that fails to materialize or if it is seen to be bolstering enemies. And famines are slow-onset disasters in places where drought, poverty, illness, malnourishment and lack of sanitation are endemic. When a crisis call is unpopular, difficult and fraught with political implications, it has to be correct. That's where the famine warning network comes in.

From the beginning, the network relied on a mix of social and physical science data to determine, more precisely than ever before, which parts of the population, in which regions of which countries, would suffer most from environmental shocks, usually drought.

The FEWS decision-support system "can be seen as an interactive filtering process by which enormous amounts of data are transformed into fair, objective, reproducible and defensible analyses," wrote two of its scientific supporters, Chris Funk and James Verdin from the U.S. Geological Survey (USGS). "Effective early warning combines a successful blend of Earth observations, hydrologic modeling, food economics, weather and climate modeling, and much more."[8]

While the network does warehouse data in computerized environments, it does not analyze all of them using automated analytical algorithms, preferring instead to merge human and software analysis. "Theoretically it could be computerized," Eilerts says. "We have had people offer to organize the data, but we're afraid that the overhead and the outcomes aren't worth the risk of disrupting our activities."

What's more, he isn't sure software could capture and make sense of the subtleties and nuances that long-experienced observers can apply. For example, he says, all the science and much of the social data may show that the food crisis is approaching famine level in Nigeria. "But I know the area, and a lot of people are sending remittances home from outside Nigeria, so it's not," Eilerts says. "Each situation is very local; it's about human behavior."

The USGS is the network's most active science partner, but it's hardly alone. Even Eilerts, who is housed in USAID, is a USDA employee. USAID built the famine network through interagency agreements with USGS, USDA, the National Aeronautics and Space Administration (NASA) and the National Oceanic and Atmospheric Administration (NOAA). The science agencies provide satellite remote-sensing data, modeling, forecasting, geographic information systems, training and analysis.

"We focused very early on evidence and data and science," Eilerts said. "We didn't flail around. It was, 'How do we get data and information to say if there is a [food] access, availability or utilization problem.'"

Network food security analysts in Africa, Central America and Afghanistan track market, vulnerability, livelihood and agricultural conditions. A contractor compiles the analysis into tightly written food security outlooks, alerts and briefs for decision-makers in USAID's Food for Peace and disaster assistance offices, the State Department, Congress, the White House, the United Nations World Food Program, humanitarian assistance organizations and other governments.

Combining physical and social science takes close interaction, wrote NASA research scientist Molly Brown. "For example, an analysis of the impact of drought as measured both by vegetation anomalies and rainfall deficits needs to be integrated with information on elevated food prices, migration patterns and water scarcity."[9]

Research and development agencies, such as NASA, look for opportunities to get their new techniques for using satellite data and modeling put into use, and the famine network provides one. "We help them see ways in which their science and technology can be effectively applied for an activity like FEWS NET, which is quite operational in nature," said Verdin.

"Same thing with NOAA," he added. "The weather service is very operational. We help them by contributing land surface monitoring products to the atmospheric, which is their bread and butter."

The agencies now use data they developed for famine warning to analyze conditions elsewhere, according to Verdin. "Our expertise with the vegetation index has helped us to monitor indications of drought and fire hazard in the U.S," he said. "Another example you've got later is the land surface temperature data. We first used it to help USAID in Afghanistan and now use it in regional offices."

The opportunity for scientific communication, rather than authority over one another, sustains the network—that and "just a common desire to do the best job we can do by FEWS," he said.

NASA is tailoring a version of its land information system (LIS) just for FEWS NET. The software framework contains several different land surface models that each help characterize a continent or country in terms of soil, vegetation coverage and other land attributes. "It's a way to transfer your measurements into land surface conditions—the depth of rivers, snowpack, etc.," Verdin said. "The advantage of having an LIS, [is that] if

8 Chris Funk and Jim Verdin, "Real-time Decision Support Systems: The Famine Early Warning Systems Network," *Satellite Rainfall Applications for Surface Hydrology*, eds. Mekonnen Gebremichael and Faisal Hossain, (Springer Science+Business Media B.V., 2009).

9 Molly Brown, *Famine Early Warning Systems and Remote Sensing Data* (Springer-Verlag Berlin Heidelberg, 2008), Kindle location 595.

each of the models is a little better or worse at doing different things … it's an ensemble, and it's a better answer than just doing one."

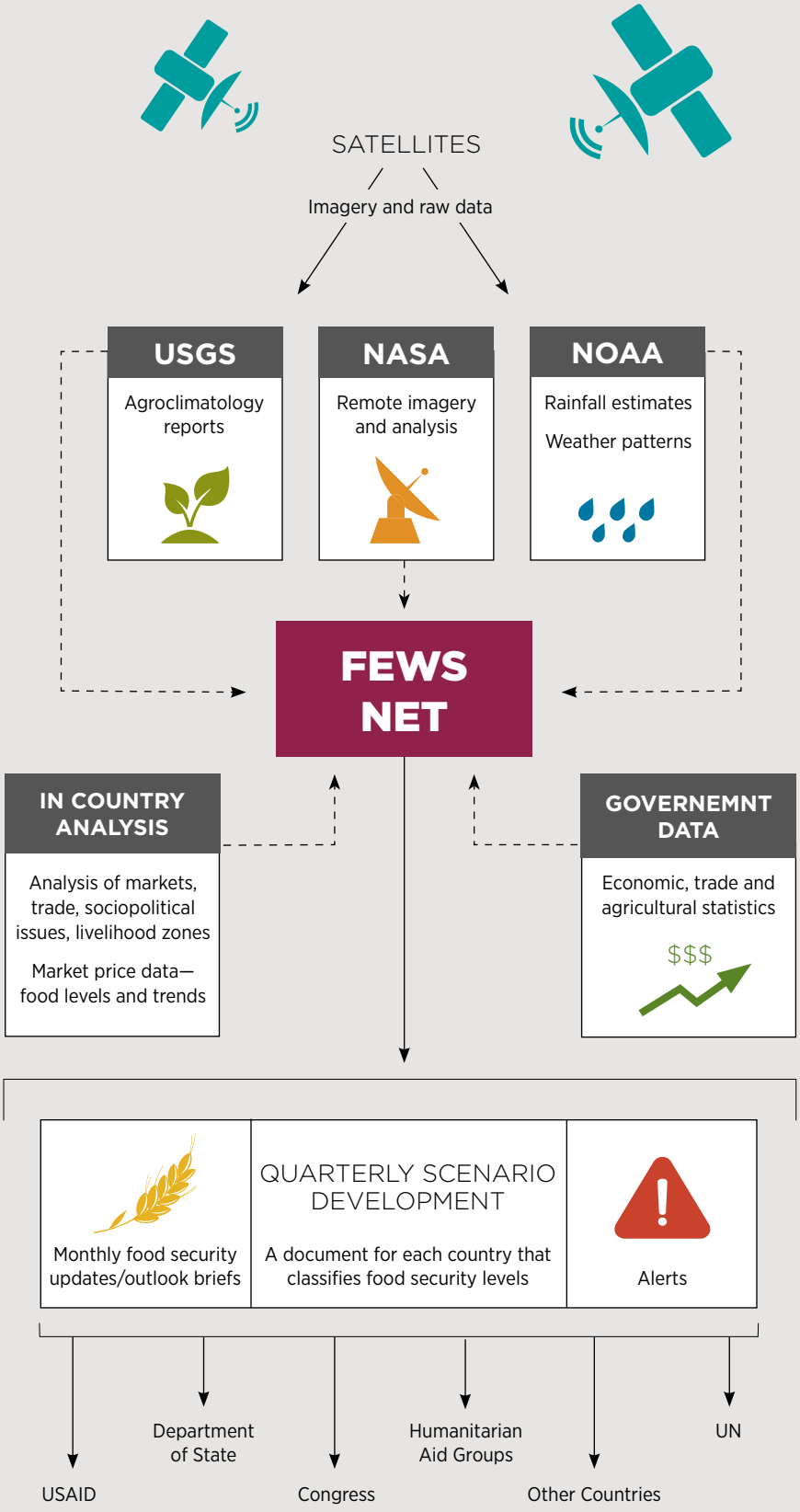
Like FEWS NET, the land information system is collaborative. Created in 2002 by NASA with help from Princeton University, the nonprofit Center for Oceans, Land and Atmosphere, and NOAA's National Centers for Environmental Prediction, LIS proved so useful that others began to fund further development, according to the framework's initial developer, NASA scientist Christa Peters-Lidard. Beginning in 2006, the Air Force Weather Agency (AFWA) entered into a reimbursable agreement with NASA. The Army Corps of Engineers also ponied up.

The Air Force traditionally has been the lead weather service for the Defense Department (DOD). Land-based branches of the military, such as the Army Corps of Engineers, rely on the AFWA to predict how soil moisture will affect movements of heavy military equipment wherever they are deployed. The land information system makes those forecasts more precise.

It also helps the weather agency feed DOD's tactical decision aid system, which combines environmental data with target and background characteristics, celestial information, angle of attack and other data to determine which weapons systems are best for a mission.

"The successes that we've had within Air Force and Army and NOAA and NASA all working together with this system, I think it is a message to be carried that not all of us in government are out to create duplicative efforts," said John Eylander, who was the weather agency's chief technologist when he began collaborating with NASA's Peters-Lidard after they met at a 2005 conference.

"I think that our successes are not necessarily just because of LIS. I think the real benefit we have is that we have a group of individuals that are really interested in working together," added Eylander, who now works at the Army Corps of Engineers.



CDC and State Health Teams Use DNA Fingerprints to Collar Bad Bacteria

PulseNet synthesizes, coordinates laboratory data to detect national foodborne outbreaks

Between November 1992 and February 1993, four children died and 732 people became ill after eating E. coli–contaminated hamburgers served at Jack in the Box restaurants in Washington, Idaho, California and Nevada. The outbreak inspired a national effort to speed detection of foodborne illnesses.

In 1994, the Centers for Disease Control and Prevention (CDC) and the Association of Public Health Laboratories (APHL) began work on a database containing the DNA fingerprints of the E. coli bacterium. By 1996, the database, known as PulseNet—the same name as the network of state laboratories working with CDC—was up and running, processing 154 bacterial DNA samples and identifying several multistate outbreaks in its first year. Since then, PulseNet has accumulated more than 500,000 bacteria isolates.

Eighty-seven public health laboratories, including at least one in every state, participate in PulseNet, creating partial images of bacterial DNA using equipment purchased with CDC grants. Fecal, blood or urine samples from sick patients are sent from hospitals and doctors' offices to local labs, which extract cultures of bacteria and send them to the public health labs for DNA fingerprinting. The prints then go to PulseNet for analysis.

The CDC requires all the PulseNet laboratories to use the same data standards, ensuring that bacteria strains can be compared in a single, shared database. PulseNet scientists run these images against the database for matches with samples from other patients. More than two matches form a "cluster," representing a possible outbreak.

When CDC analysts discover a cluster that is larger, faster-forming or more dangerous than is typically expected during that time of year, they alert state public health epidemiologists to investigate. Interviewing patients about their recent food handling, exposure and ingestion helps state health officials identify the source of the bacteria and stop the outbreak.

The CDC estimates that 47.8 million people a year get foodborne illnesses, resulting in 127,839 hospitalizations and more than 3,000 deaths.[10] The annual economic burden of foodborne illness ranges from $51 billion to more than $77 billion.[11]

In its first year, PulseNet interacted with four public health laboratories and tracked a single pathogen. By 2012, PulseNet included 87 labs and was tracking eight pathogens. PulseNet-certified Food and Drug Administration (FDA) and Agriculture Department (USDA) laboratories also use PulseNet to track pathogens collected from food or animals in an attempt to catch illness-causing bacteria earlier, before they infect people.

These partnerships flourish because CDC meets regularly with FDA and USDA to discuss data standards. CDC branch chief Ian Williams reflected: "Like any good relationship or marriage, it requires working together to identify and resolve problems as they come up, and we're good at that."

In 2011, officials traced Listeria to cantaloupes from a farm in Colorado. It was the deadliest foodborne disease outbreak in the United States in almost 90 years—causing 29 deaths and a miscarriage. In just 10 days, officials spotted an unusual increase in Listeria cases in local hospitals, identified contaminated cantaloupes as the source and issued a national consumer warning. Officials said it was "the fastest Listeria investigation [they'd] ever seen."[12] "Up to twice as many would have been infected had officials not had the tools, people and systems in place," estimated CDC deputy director Robert Tauxe.[13]

Food is not the only bacteria carrier PulseNet can trace. In 2012, PulseNet discovered a Salmonella outbreak; 26 people in 12 states were infected. Eight were hospitalized and one died. Epidemiological investigations led to pet hedgehogs.

Ohio State University's Robert Scharff estimated that PulseNet costs about $10 million a year and saves $291 million.[14]

The CDC pointed out that during the 1997 E. coli outbreak in Colorado, "If 15 cases were averted by the recall of potentially contaminated ground beef, the PulseNet system [in that state] would have recovered all costs of start-up and five years of operation."[15]

In an era of continuing fiscal uncertainty, PulseNet must continue to demonstrate its value. A risk to PulseNet's effectiveness is the rise of very fast and inexpensive clinical laboratory tests that do not produce a pure bacteria culture. Without bacteria to "fingerprint," PulseNet can't track outbreaks.

So CDC plans to ask labs to continue submitting cultures and preserve "fingerprints" of samples to help halt future outbreaks. Meanwhile, CDC is working on developing sequencing technology for genetic material that doesn't require a pure culture of bacteria.

Whole genome sequencing—mapping an entire strand of DNA—offers PulseNet a tantalizing opportunity. Genome sequencing produces more DNA data than the current testing, which would increase the speed and accuracy of PulseNet's bacteria identification. As sequencing becomes further automated and problems with transmitting and storing its large images are solved, it will allow CDC to include in PulseNet's database all known pathogens instead of the eight currently tracked.

10 CDC, 2011 Estimates of Foodborne Illness in the United States, http://1.usa.gov/17iULmd.

11 Robert L. Scharff, "Economic Burden from Health Losses Due to Foodborne Illness in the United States," *Journal of Food Protection*, Vol. 75 No. 1, 2012, 123–131.

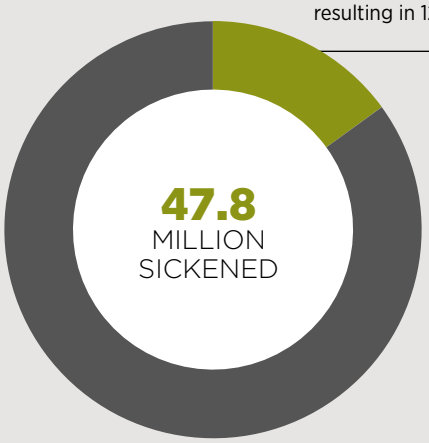
12 Quoting Robert Tauxe, MD, MPH, deputy director of CDC's Division of Foodborne, Waterborne and Environmental Diseases, from "Deadly Listeria Halted in Record Time," http://1.usa.gov/17OlNfl.

13 Tauxe, "Deadly Listeria Halted in Record Time," available http://1.usa.gov/17OlNfl.

14 Robert L. Scharff, "A Model of Economic Benefits and Costs from PulseNet," The Ohio State University Department of Consumer Sciences, 2010, presentation slides, http://bit.ly/18GMGmU.

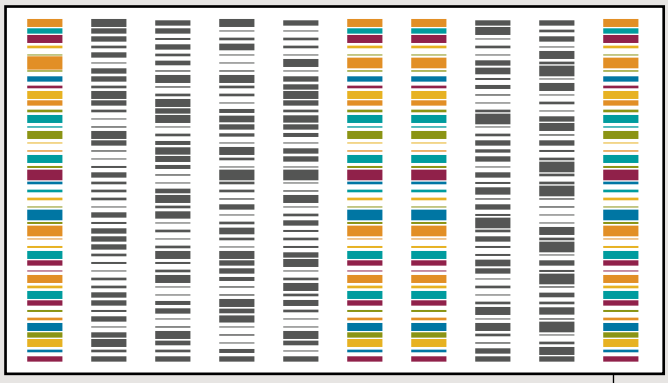
15 EH Elbasha, TD Fitzsimmons, and MI Meltzer, "Costs and Benefits of a Subtype-specific Surveillance system for Identifying Escherichia coli O157:H7 Outbreaks," Emerging Infectious Diseases, 2000, Vol. 6 No. 3, 293–297.

The CDC estimates that 47.8 million people in the United States (15%) fall ill due to foodborne illness annually, resulting in 127,839 hospitalizations and 3,037 deaths.



Of those who fall ill, fewer than 1 in 15 seek medical care.

If a doctor suspects the illness to be foodborne in nature, a sample is sent to a local clinical lab for testing. The result is sent back to the doctor for follow-up patient care and, if a pathogen is found, the bacterial sample is sent to a public-health laboratory for further diagnosis and subtyping.



PulseNet public-health laboratories create DNA "fingerprints" of the bacterial sample and compare these with other samples in the database. Two or more matches represent a "cluster," indicating a possible outbreak.

200
CLUSTERS
IDENTIFIED

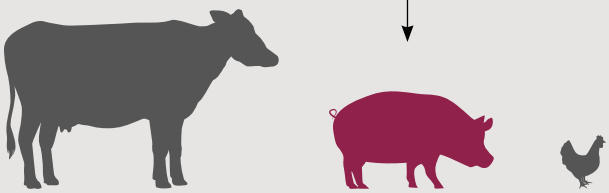
If a cluster is atypical, state and local epidemiologists work to identify the source of the outbreak. CDC may coordinate efforts involving multiple states.

167
INVESTIGATIONS
PERFORMED

IN 2012

60
SOURCES
IDENTIFIED

9
RECALLS
PERFORMED



Once the source is identified, the CDC works closely with FDA and USDA regulators to stop the outbreak, determine when the outbreak is over and prevent future outbreaks.

Biometrics Nets Bad Actors in Afghanistan, Iraq

Despite successes, program lacks a strong sponsor, training and an official home

Afghanistan government forces—primarily police and military—have been killing a larger percentage of soldiers from the U.S.-led coalition each year. The percentage of so-called green-on-blue killings grew from less than 1 percent in 2008 to 15 percent of all coalition deaths in 2012. The 80 attacks between January 1, 2008 and July 15, 2013, left 134 coalition troops dead and 153 wounded.[16]

So when biometrics helped prevent a man on a U.S. terrorist watch list from joining the Afghan police this June, Don Salo, director of the Defense Department Forensics and Biometrics Agency, counted it as a victory. "You can imagine the implications for avoiding potential green-on-blue attacks [of] catching the bad guys before they get approved by the Afghan army or local police from gaining access to our bases," he said during a June biometrics symposium.[17]

By 2012, there had been at least 3,000 matches on biometrics collected from 1.2 million non-U.S. persons in Afghanistan, according to the Government Accountability Office (GAO).[18]

Including biometrics collected in Iraq beginning in 2002 through December 2011, the U.S. military has created more than 7.1 million records—sets of facial photographs, iris scans and 10 fingerprints—from 4.5 million people. Between fiscal years 2007 and 2012, DOD invested $3 billion in the program.[19]

"Over the past decade, the U.S. Army grew a program that led to the successful targeting of over 850 high-value individuals, denied access to over 64,000 potential threats and resulted in over 200 interdictions through collaboration with interagency partners," said John Boyd, Pentagon biometrics director. "Leveraging biometrics enables ongoing operations and translates to real successes for the security of our nation."

Yet defense biometrics analysis has been criticized for lacking a single strong sponsor within the Pentagon, insufficiently training collectors and their commanders, buying mismatched collection devices, and not fixing slow transmissions to and from the central database—the Automated Biometrics Identification System (ABIS).

Begun when war funding for Iraq and Afghanistan was flowing freely, the biometrics program is on shaky ground now that U.S. troops have left Iraq and will leave Afghanistan after next year. The special pot of money for overseas contingency operations is dwindling as deployed U.S. forces come home.

The Army holds the responsibility for biometrics for all the military services, but authority has been split among eight Army organizations. "A lack of collaboration between [Army] components has led to direct failures," according to a 2012 report by the RAND National Defense Research Institute.[20]

The program was whipped up quickly in reaction to soldiers' urgent needs on the frontlines in Iraq. The Army never made it a program of record, instead purchasing biometrics collection devices as quick-response capabilities.

This led to a proliferation of collection equipment—more than 7,000 of three different types of devices have been fielded, much of it not interoperable because the Army rapidly purchased whatever companies had available.

"The Army has been forced to respond to urgent operational needs from Iraq and Afghanistan, which resulted in tools and technology being rapidly developed and fielded without adhering to DOD standards, formal performance measures and operational testing and evaluation requirements," RAND found.[21]

Not being a program of record kept biometrics from being taught in Army schoolhouses, so pre-deployment training is ad hoc and troops often aren't proficient at capturing full fingerprints or identifying which hand or person the prints are from.

If service members fail to collect biographical information from subjects or it is lost, the data become useless, "such as the approximately 4,000 biometrics collected from 2004 to 2008 that were separated from their associated identities," according to GAO.[22] When money was no object, the services could pay contractors to collect the data, but that didn't build a DOD knowledge base, RAND noted.

When there's a match between biometrics collected in the field and a person on a watch list, the information goes to analysts and intelligence units, but often not to the troops who collected the data.

That reduces the enthusiasm of the men and women on the ground doing the collecting but not hearing the success stories, the "guys that go into a bad-actor house in the foothills of Afghanistan and there's 10 bad guys there that they've got to enroll before mortars start coming in," said Boyd.

"If we prevented a bad actor from entering the country … on the information and collaboration from Afghanistan, we need to feed that back," Boyd said. "If a warfighter sees no value in what he's doing, eventually he'll stop doing it."[23]

The training that troops do receive predeployment and in the field doesn't extend to their leaders, GAO found. Unit commanders aren't taught how to use biometrics effectively, get their soldiers trained to collect the data or find troops who already have those skills. But even when commanders effectively deploy biometrics in the field, their support can't overcome lack of knowledge in the top ranks.

"We have a lot of people at the major and lieutenant colonel level who get it, who saw biometrics achieve great success in theater. But some of the senior civilians and [senior people] in uniform may not appreciate the

16 "Green-on-Blue Attacks in Afghanistan: The Data," The Long War Journal, August 23, 2013, http://bit.ly/19m3HAV.

17 Biometrics, Analytics and Big Data Symposium, Technical Training Corporation, June 18, 2013, Rosslyn, Va.

18 Government Accountability Office, *Defense Biometrics: Additional Training for Leaders and More Timely Transmission of Data Could Enhance the Use of Biometrics in Afghanistan* (Washington, DC: GAO-12-442, April 2012), http://1.usa.gov/15Gteoi.

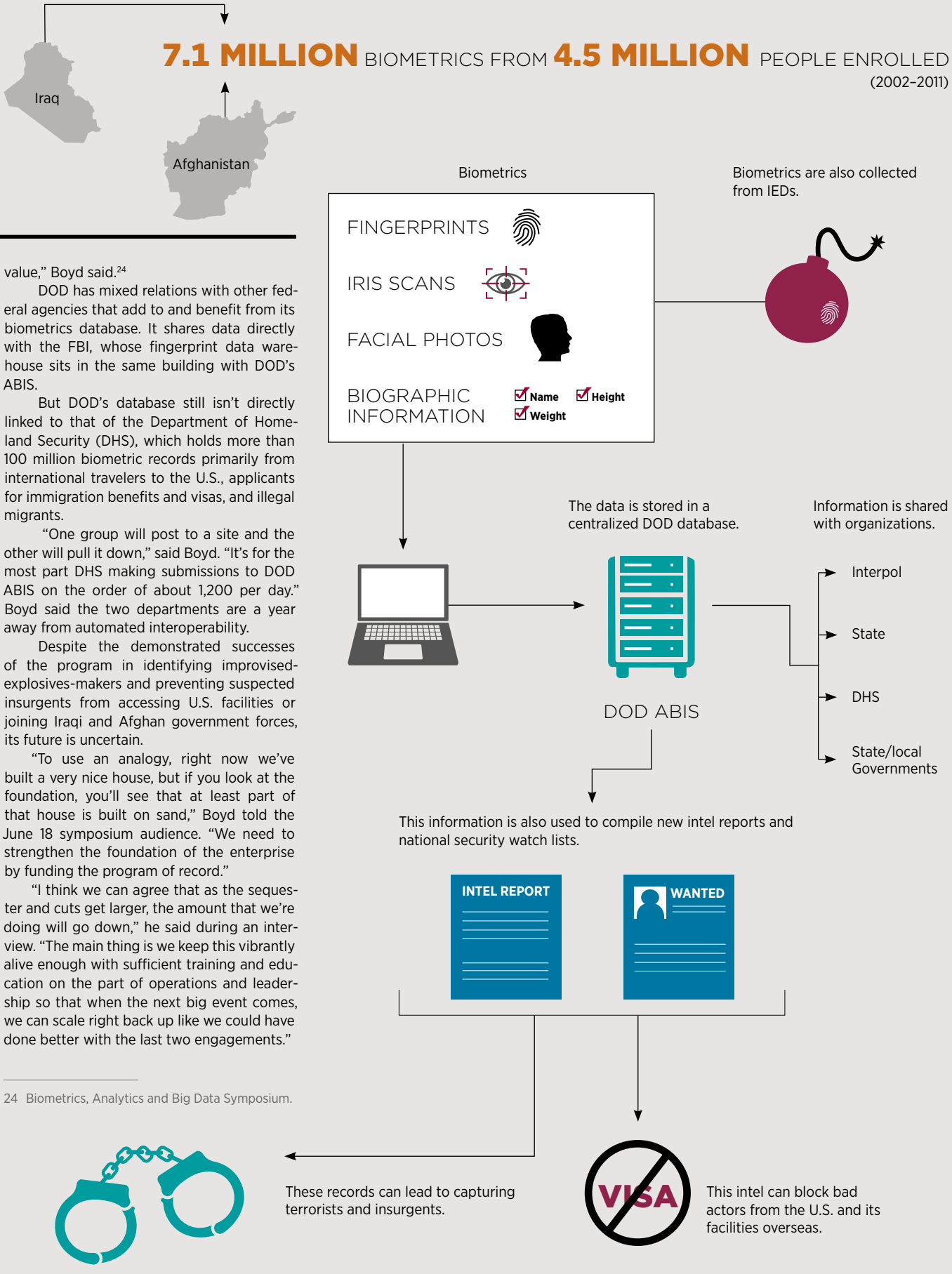
19 Douglas Shontz, Martin C. Libicki, Rena Rudavzky and Melissa A. Bradley, *An Assessment of the Assignments and Arrangements of the Executive Agent for DoD Biometrics and Status Report on the Biometrics Enterprise* (RAND National Defense Research Institute, 2012), 70.

20 Ibid., 24.

21 Ibid., 25.

22 GAO, 2012.

23 Biometrics, Analytics and Big Data Symposium.



value," Boyd said.[24]

DOD has mixed relations with other federal agencies that add to and benefit from its biometrics database. It shares data directly with the FBI, whose fingerprint data warehouse sits in the same building with DOD's ABIS.

But DOD's database still isn't directly linked to that of the Department of Homeland Security (DHS), which holds more than 100 million biometric records primarily from international travelers to the U.S., applicants for immigration benefits and visas, and illegal migrants.

"One group will post to a site and the other will pull it down," said Boyd. "It's for the most part DHS making submissions to DOD ABIS on the order of about 1,200 per day." Boyd said the two departments are a year away from automated interoperability.

Despite the demonstrated successes of the program in identifying improvised-explosives-makers and preventing suspected insurgents from accessing U.S. facilities or joining Iraqi and Afghan government forces, its future is uncertain.

"To use an analogy, right now we've built a very nice house, but if you look at the foundation, you'll see that at least part of that house is built on sand," Boyd told the June 18 symposium audience. "We need to strengthen the foundation of the enterprise by funding the program of record."

"I think we can agree that as the sequester and cuts get larger, the amount that we're doing will go down," he said during an interview. "The main thing is we keep this vibrantly alive enough with sufficient training and education on the part of operations and leadership so that when the next big event comes, we can scale right back up like we could have done better with the last two engagements."

24 Biometrics, Analytics and Big Data Symposium.

Analytics Improves Aim of Agricultural Pest Hunters

Animal Plant Health Inspection Service's risk-based system targets shipping containers most likely to carry pests

In 1909, the long-awaited shipment of 2,000 Japanese cherry trees arrived at United States ports, a gift symbolizing the friendship between the two nations. The U.S. Department of Agriculture (USDA) found non-native pests, including white peach scale, black thrips, clearwing moths and San Jose scale in the trees and had to destroy them.[25] The USDA worked with Japan to treat a new batch of trees, which arrived pest free in 1912 and continue to blossom, drawing tourists from all over the world each spring.

A hundred years later, more than a billion plant imports come into the country each year, and USDA's Animal Plant Health Inspection Service (APHIS) Plant Protection and Quarantine Division uses data analytics to inspect shipments, detect pests and prevent them from spreading to U.S. agriculture.[26]

As shipments arrive at U.S. ports, APHIS officials oversee inspections by Department of Homeland Security (DHS) Customs and Border Protection (CBP) officials. By analyzing data about the type and date of shipment and country of origin, the inspection service can target inspections to containers most likely to contain pests.

If inspectors find pests, APHIS can destroy the shipment, send it back to the exporting country or treat it and send it to market. APHIS officials also set traps around ports to catch pests that escape detection. APHIS intercepted more than 565,000 insects in baggage and cargo between 1984 and 2002, though it always misses some.[27] Accordingly, the agency works with state agricultural departments and universities to conduct delimiting surveys, which determine the scope of existing infestations in the United States.

The Government Accountability Office estimates that non-native pests cause about $136 billion in lost agricultural revenue annually. The Asian longhorned beetle, for example, arrived from China on a wood shipping pallet and since 1985 has destroyed more than 80,000 ash, maple and other trees it has bored holes into. It has caused $269 million in damages in five states, an amount that could exceed $41 billion if the beetle spreads to the entire country.[28]

Another pest suspected of hitching a ride from Asia was discovered in Michigan in 2002. The emerald ash borer has killed some 58 million ash trees in more than 10 states, causing tens of millions of dollars in damage.[29]

APHIS helps prevent pest damage using two databases: the Agricultural Quarantine Activity System (AQAS) and Integrated Plant Health Information System. The quarantine system helps manage shipment data and target inspections; the plant health information system manages information from infestations already in the country.

CBP inspectors feed their findings into the quarantine system and also are guided by it. The risk-based system, adopted in 2007, provides data that can be used in probability distributions to aid in determining how many of which crates to target, based on known pest or disease outbreaks in other countries, weather patterns and other information.[30]

With the quarantine system, the animal and plant agency can better allocate inspection resources, improving its find rates at U.S. ports and its trapping rates.

The 2,360 CBP officials who use the quarantine system for inspections at 167 U.S. ports (out of 329) do so under a 2003 memorandum of understanding between DHS and USDA. It gives the inspection agency the job of maintaining the quarantine system and training CBP inspectors, who check containers and collect data.[31]

CBP officials send their findings to the quarantine system, where it is incorporated in the risk analysis that guides future inspections. They send pests to the APHIS Plant Protection Quarantine National Inspection Service for final confirmation, information that also goes into the quarantine system.

Todd Schroeder, director of business systems at the inspection agency said, "The whole idea here [is] to protect American agriculture, you've got to be able to make risk-based decisions about imports of [agricultural] products."

Improved decision-making has increased the inspection system's find rates and productivity and changed its approach to core mission activities. For instance, by knowing which shipments may have what problems, APHIS can turn its attention to working with offshore partners to prevent pests from boarding shipments in the first place.

APHIS provided 1 million classical swine fever vaccines to Guatemala in March, for example, to prevent the disease from spreading. "With trade opportunities increasing each year, the promotion of animal health across borders is important now more than ever," said John Clifford, USDA's chief veterinarian. "APHIS recognizes that the prevalence of animal disease in one country could easily transpose to another."[32]

Analytics could improve the inspection system's pest detection to the point that it could invest fewer resources in inspection and turn toward prevention, working with food companies and other countries, said Schroeder. "You may be making decisions that impact different industries in different ways," he said. "You have less resources to spread across different programs, so let's focus our resources on those areas [where] we can have the largest overall impact."

25 USDA, *The Japanese Flowering Cherry Trees of Washington, D.C.: A Living Symbol of Friendship;* National Arboretum Contribution, No. 4, http://1.usa.gov/1bczcAg.

26 USDA, *APHIS' Plant Inspection Stations: Protecting American Agriculture from Foreign Pests and Diseases*, APHIS Program Aid No. 1942 (2007), http://1.usa.gov/1adFJwm.

27 "The Emerald Ash Borer Facts," USDA (2009), http://1.usa.gov/1ccrk76.

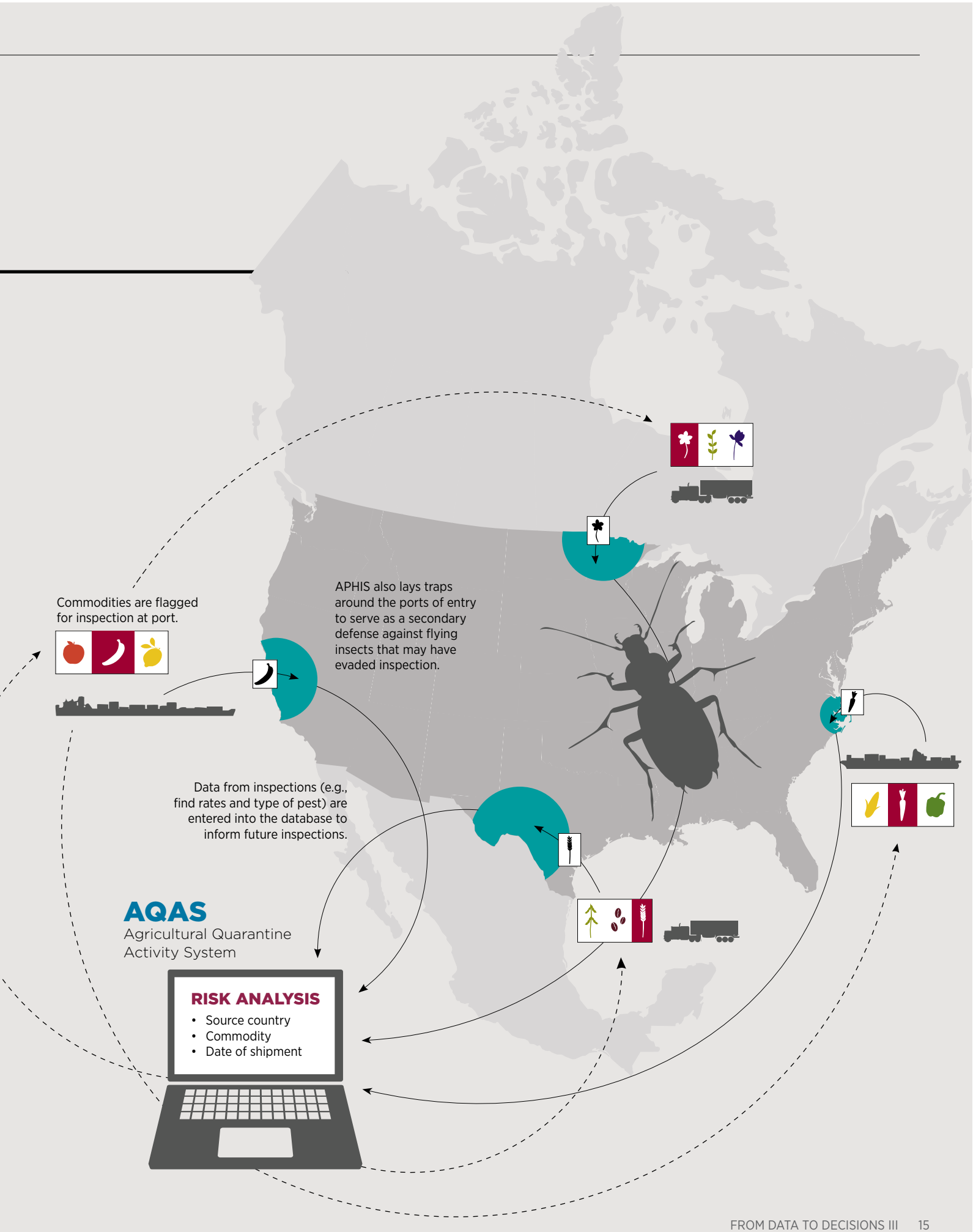
28 GAO *Agriculture Inspection Program Has Made Some Improvements, but Management Challenges Persist*, (Washington, DC, GAO-12-885, September 2012), http://1.usa.gov/1a5JqGf.

29 Ibid., USDA

30 USDA, Questions and Answers: New Risk-Based Sampling Protocol and Propagative Monitoring and Release Program at Plant Inspection Stations Factsheet Jan. 2012, http://1.usa.gov/179nrvD.

31 Memorandum of Agreement between the United States Department of Homeland Security (DHS) and the United States Department of Agriculture (USDA), DHS Agreement Number: BTS-03-001, USDA-APHIS Agreement Number 01-1001-0382-MU, http://1.usa.gov/HgpjZ1.

32 "USDA Donates One Million Doses of Classical Swine Fever Vaccine to Guatemala," APHIS News Release, Washington, March 25, 2013, http://1.usa.gov/1a5JK7.


Commodities are flagged for inspection at port.
APHIS also lays traps around the ports of entry to serve as a secondary defense against flying insects that may have evaded inspection.
Data from inspections (e.g., find rates and type of pest) are entered into the database to inform future inspections.
AQAS
Agricultural Quarantine Activity System
RISK ANALYSIS
• Source country
• Commodity
• Date of shipment

Taking Veterans' Care from Random to Routine

The Veterans Health Administration builds a warehouse and many medical tools from its homemade electronic health record

Like most primary care physicians, Stephan Fihn is always busy, regularly behind and often frazzled.

Left to his own devices, when seeing a patient whose age and chronic illness might qualify him for home health care, Fihn was lucky if he even remembered that the Veterans Affairs Department's Puget Sound Healthcare System offered it.

"A patient would come in every three to four months to the clinic, and if I wasn't too distracted, if I wasn't too overwhelmingly busy, and not too behind on my schedule, I would say maybe this patient should go to home care," Fihn recalled. "So whether that patient got into home care depended whether I thought about it, or whether they asked for it, or whether I was familiar with the program. It largely bordered on being a random event."

But home care keeps patients out of the hospital, and can keep them happier, healthier and alive longer, while saving the VHA money. So the VHA needs its primary care providers to get eligible patients enrolled.

Fortunately, Fihn directs VHA's Office of Analytics and Business Intelligence, positioning him to do something to make doctors' referrals routine.

First, Fihn helped come up with the care assessments needs (CAN) score, a predictive analytic tool intended to identify which of VHA's 6.5 million primary care patients are at highest risk of hospitalization or death.

Now his office is creating an online care coordination tool, the patient care assessment system (PCAS), which uses these assessment scores and hundreds of other data points to help 7,000 VHA medical teams coordinate 900 to 1,200 patients each.

"We have a [registered nurse] care coordinator. She's got this list in front of her with patients who are at highest risk and she can go down it systematically now and say, 'Oh I see Mr. So-and-So was in the hospital twice in the last month and he's got a really high risk score and we're not giving him any of these services. Maybe I will go talk with his primary care provider and figure out in our daily huddle which of his patients should be referred to one of these programs,'" Fihn said.

"It's really changed the paradigm from a reactive one, where maybe in the end the only way we would see we were not doing it well was [high] hospitalization rates and we'd have to figure out why," he said. "Now we're looking at our patients at high risk of being in the hospital and saying, 'What services can we provide to keep them out?'"

Both the scores and the assessment system draw from the 80 terabytes of patient statistics and other information collected in VHA's corporate data warehouse, which sucks up data from electronic health records (EHRs) stored in the Veterans Health Information Systems and Technology Architecture (VISTA) at VHA medical facilities. VISTA was created 20 years ago by rogue software tinkerers, many of them doctors, who saw in early personal computers an opportunity to improve how they managed and cared for patients.

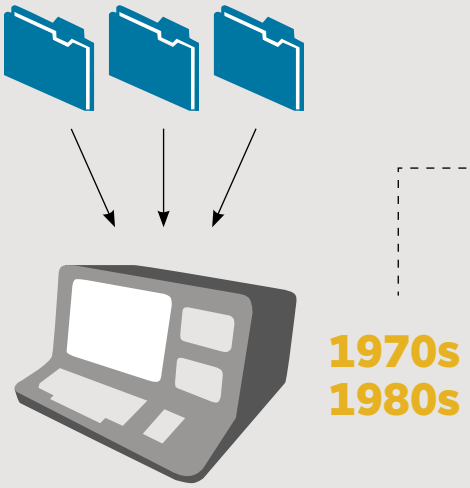
The software they designed evolved into VISTA, which VHA offers free online. The EHR system is one of the most widely used in the world. The presence of all that data has inspired VHA analysts to continue innovating and building tools to use it. The data warehouse made the assessment scores and system possible, and helped VHA "do greater comparisons of our patients as they move across our system and as we see differences in treatment," said Gail Graham, the VHA's deputy undersecretary for health for informatics and analytics, and Fihn's boss.

She credits him with building the culture that makes data users out of VHA staffers at all levels. Fihn "is really growing analytics from a basic level of making sure that there is a cadre of people in medical centers and clinics that know how to use Excel, all the way to offering university-level courses in using the data and advanced analysis."

As a result, VHA employees refine and invent new uses for informatics products. The PCAS, for example, was built for registered nurses in concert with registered nurses, who coordinate care for most teams, says clinical program manager Joanne Shear. "Dr. Fihn's office heard the hue and cry across the country: 'How do we know exactly who we have to manage, how do we identify them? Where's the tool that we can interact with the medical record?'"

System developers fully expect the assessment tool will morph as nurses use it—but only if they see it as a help, not a hindrance, says VHA health information technology lead Tami Box. "The best and most optimized use remains be seen. We are trying to put tools in the hands of the people to use them and then find out how they use them in the best ways."

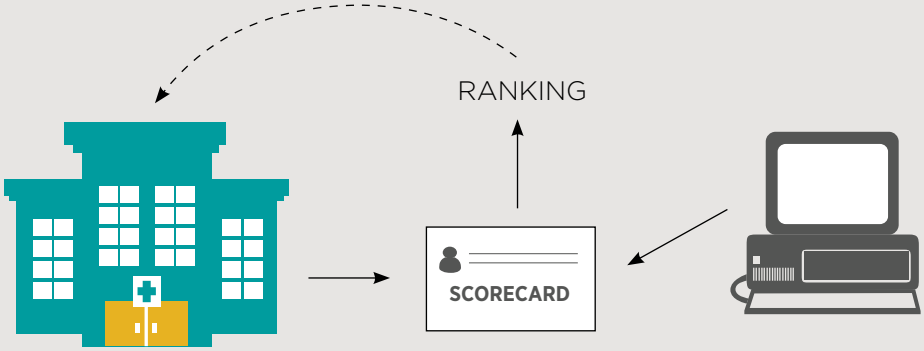
Box and her team take pains to ask little of first-time users so as not to interrupt their delivery of care. The first release of PCAS, now in use across VHA hospitals, "doesn't have a whole lot of places where users have to enter data," she adds.

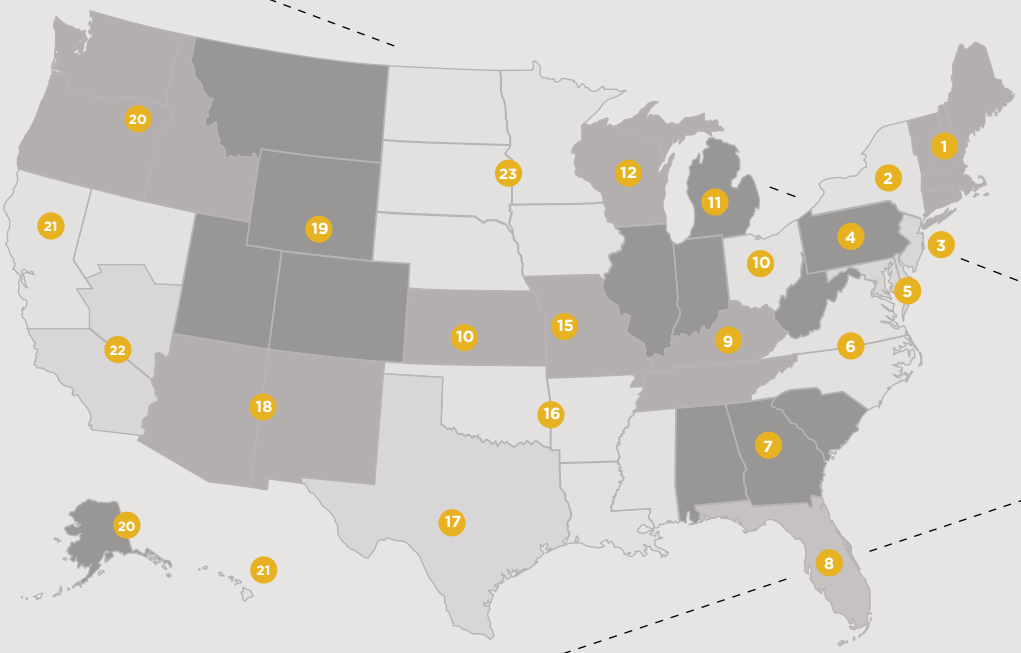


Tinkering doctors start using minicomputers to store patient files. After initial resistance, the homemade electronic health records are adopted VHA-wide.

1990s

Electronic health records enable VHA to measure its quality of health care against U.S. standards and improve.





1994

VHA is decentralized to increase autonomy and improve regional care, forming 22 (now 21) Veterans Integrated Service Networks (VISNs), comprising VA hospitals, health care centers, ambulatory care centers and community-based outpatient clinics.

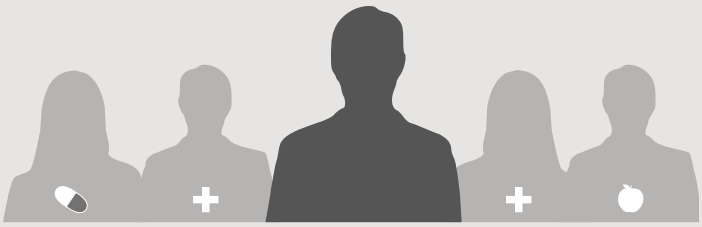
2003

VHA launches MyHealtheVet, providing patients with remote access to their electronic Personal Health Record.

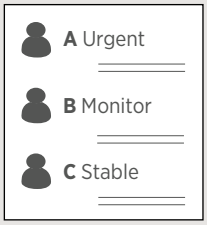
VHA begins the Care Coordination/Home Telehealth (CCHT) program. Telehealth devices include digital cameras, videophones, messaging devices and biometric and telemonitoring devices.

EXPANDED TELEHEALTH

2010



VHA begins a patient-centered medical home model that organizes primary care around interdisciplinary Patient Aligned Care Teams (PACTs), each with a primary care provider, registered nurse, licensed practical nurse and medical clerk who share responsibility for managing a group of nearly 1,200 patients.



2011

VHA introduces care assessment needs (CAN) scores to help care teams prioritize services to high-risk patients. In 2012 the CAN score becomes a part of the new care scheduling system for the teams.

Collaborate with other agencies to collect data and share analytics expertise

Most analytics pathbreakers are strikingly collaborative. They seek out like-minded souls who might already have collected data they can adapt to their purposes or who have developed new methods for prying out or combining it.

The U.S. Agency for International Development (USAID), for example, specifically created the Famine Early Warning Systems Network (FEWS NET) to be collaborative. From the start, USAID paid other agencies for expertise, data and analysis. The network comprises staff from the Agriculture Department (USDA), scientists and data from the U.S. Geological Survey (USGS), analysts and data from the National Aeronautics and Space Administration (NASA) and the National Oceanic and Atmospheric Administration (NOAA), contractor decision support and in-country staff, as well as alliances with a host of humanitarian aid and government agencies.

Most famines are caused by environmental shocks, such as droughts. USAID knew plenty about food assistance, but far less about the causes of those shocks and the socioeconomic conditions that turn them into food emergencies. To find out, the network draws data from a variety of sources, including meteorology and climatology, agricultural monitoring and harvest assessment, food market and trade analysis, health and nutrition outcomes, livelihoods analysis and food-needs tracking.

From its beginning in 1986, FEWS NET used data from social and physical scientists to improve how USAID targets the most vulnerable populations for food aid. For more than 15 years, it has used participating agency program agreements (PAPA) and participating agency service agreements (PASA) to pay other agencies for people and information.

"Somebody would say, 'I understand these guys have a new way of measuring rainfall, let's go talk with them,' and they would say, 'There's this guy over at NASA that actually is in control of the imagery, so why don't you make an arrangement with him,'" said FEWS NET program manager Gary Eilerts. "So we made an arrangement with [NASA's] Goddard Space Flight Center to get imagery ... then we had [NOAA] giving us better rainfall estimates ... and USGS trying to help us put it together in spatial perspective."

Eilerts, a USDA employee who works at USAID managing the famine network, draws up and oversees agreements with NASA, NOAA, USGS and USDA. The agreements permit USAID to pay other agencies for inherently

governmental services and facilities that are "particularly or uniquely suitable for technical assistance, are not competitive with private enterprise, and can be made available without interfering unduly with domestic programs."[33]

The famine network has a program agreement with the geological survey for six scientists and two managers, for example, while the NASA agreement covers satellite-collected data products on vegetation, rainfall and other underlying environmental causes of food insecurity and famine, particularly climate change.

DON'T REINVENT THE WHEEL

NASA is building a version of its land information system (LIS) for the famine network. Its Goddard Space Flight Center developed the software framework to improve assessment of ground conditions. It lets scientists run more than one model of a portion of the Earth's surface at a time to come up with precise predictions of soil moisture and other interactions between the atmosphere and the land. The land system is itself a collaboration: The Air Force Weather Agency (AFWA) and Army Corps of Engineers helped pay for NASA to develop it.

"When I provide an investment into the LIS system, that's because I need a certain capability added to it," said John Eylander, a co-founder of the collaboration between NASA and the weather agency. "Everybody basically said, 'Let's not reinvent the wheel. Let's go and collaborate with NASA and benefit from that partnership.'"

The science agencies in the famine network benefit not only by receiving funds for remote sensing and data products, but also by getting feedback on those products. The network's in-country staffers compare satellite data with what they see on the ground, for example, a quality check whose importance "cannot be overstated in the development and maturation of remote sensing," wrote NASA biospheric research scientist Molly Brown.[34]

Every year, the network holds a science day to keep its social and physical scientists aware of each other's activities, discoveries and capabilities, says USGS's Jim Verdin. "We bring each other up to date through a series of presentations," he said. "It keeps us abreast of what our colleagues are doing, but it also informs the social science food analysts as well." Learning about advances in physical science data collection and analysis lets FEWS NET social scientists understand what new or more sophisticated questions they can get answered.

The meetings also uncover new data that leads to better analysis. For example, Verdin discovered at a recent meeting that NASA had funded one of the geological survey's university partners to review past satellite images to determine the amount of water accumulated seasonally in the United States and Central Asia. "It's going to give us the opportunity to go back and compare the model in real-time and retrospective analysis."

Some analytics-driven agencies rely on others not just to share, but also to collect their data. For example, the Agriculture Department's Animal Plant Health Inspection Service (APHIS) populates its Agricultural Quarantine Activity System (AQAS) with data collected by agricultural import inspectors working for the DHS's Customs and Border Protection (CBP) bureau.

The quarantine system, adopted in 2007, helps CBP determine which containers to inspect based on data about known pest or disease outbreaks in other countries, weather patterns and other information that increases the likelihood that a shipment is infested.

In 2003, DHS and USDA signed a memorandum of agreement assigning about 3,000 APHIS inspectors to CBP. Under the agreement, APHIS trains the inspectors.[35] The service also relies on state and local governments, academics and alert citizens to report unusual instances of damage to plants or other indications that bugs or diseases escaped the ports.

CDC also turns to others for data, relying on 87 public-health laboratories across the country, at least one per state, to provide samples to compare against the database of more than 500,000 bacterial DNA fingerprints in PulseNet. CDC provides grants to the labs to buy the equipment and to pay some of the employees who use it to prepare samples.

CDC scientists run the DNA images captured from samples taken from suspected victims of foodborne illness against the database to detect multistate outbreaks. Scientists analyze matches of more than two similar patterns, which are considered possible outbreaks. Confirmed results trigger investigations by local public health department epidemiologists to find the source.

In addition, Food and Drug Administration and USDA laboratories have begun using PulseNet to track data on pathogens to catch illness-causing bacteria before people get sick. USDA and FDA also have the authority to recall products or regulate meatpacking plants and farms as a result of PulseNet's and public health officials' findings. The wealth of information and collaboration

33 Famine Early Warning Systems Network U.S. Geological Survey Participating Agency Program Agreement.

34 Brown, *Famine Early Warning Systems and Remote Sensing Data*.

35 2003 Memorandum of Agreement between DHS and USDA, http://1.usa.gov/GzzeZN.

is allowing a quicker response that keeps consumers safer.

"In the past, CDC handed off the identification of a foodborne bacteria to FDA with little more information or assistance. What's happened in the last few years is bringing regulatory partners [FDA, USDA] upstream into the process," said CDC's Ian Williams. "Our lane is 'what's causing the outbreak,' and they get the food off the market, but we really collaborate in the gray area in the middle, which is becoming quite a large area."

NASA's applied sciences research and development programs, such as the LIS, must develop transitional partnerships with organizations that can put the research and development projects into practice. The leading land system scientist and her team at Goddard were completing the last couple of years of initial NASA funding when she "hit the road to socialize the LIS concept," and the AFWA showed interest, Eylander said.

As Air Force weather chief technology officer at the time, Eylander invested in the system with money from a special fund for updating, improving and delivering new weather science that stems from authority provided under the Economy Act (1932).

That act permits federal agencies to pay one another for goods and services as long as they can provide them more cheaply and conveniently than the private sector can. Eylander's current organization, the Army Corps of Engineers Cold Regions Research and Engineering Laboratory, uses the statute's acquisitions clause to sponsor collaborative research and development with other non-DOD government organizations that can accept reimbursable funds. "It's a very handy way to encourage interagency collaboration and reduce duplicative projects."

Eylander can use the clause to benefit current projects by buying science data from NASA or the AFWA, for example. "I think you get more out of the investment because the collaboration usually leads to increased productivity," he said.

"I don't have all the answers myself, but they're very, very difficult problems … So the only option you have is to share and collaborate."

INSIGHTS FOR YOUR ANALYTICS PROGRAM

Analytics pioneers shared and added to one another's data and expertise in a variety of ways:

- Most often, they used legal authorities to buy data and the experts and software to analyze it.
    - Some used the government-wide provisions for interagency acquisitions under the 1932 Economy Act.
    - Others relied on agency-specific authority, such as the participating agency program and service agreements provided for under the 1961 Foreign Assistance Act, which created USAID and permits it to use other agencies' resources when they are uniquely suitable for technical assistance in education, health, housing or agriculture.
- Another form of interagency agreement, a memorandum of understanding, enabled CBP to collect data and USDA's APHIS to analyze it, helping both agencies meet their mission goals.
- NASA has created research and development programs whose funding is contingent on recipients' promoting their products to other agencies that can apply them and might invest in developing them further.
- CDC used grant money to help public health labs acquire the equipment they use to process DNA samples for matching against the PulseNet database.
- Annual science days give FEWS NET collaborators insights into each other's work on famine, preventing duplication and augmenting other projects across participating agencies.

Develop data to demonstrate return on investment

Mature analytics programs have struggled to define and measure the outcomes of their efforts. New projects, too, are challenged to demonstrate return on their data investments. But as programs vie to survive deficit-reduction budget cuts, demonstrating ROI no longer is optional.

The most powerful ROI estimates mix real-world results and cost-benefit analysis. The Centers for Disease Control and Prevention applies such a measure for PulseNet. Its outcome measure is the number of outbreaks identified and number of people sickened by each one. The cost-benefit analysis measures the cost of the program against the health expenditures it prevents.

So far, the number of outbreaks detected is up and the number of people sickened is down since PulseNet's inception. It began in 1996 and helped identify 13 outbreaks in the first six years. In the next five years, PulseNet caught 19 outbreaks. On the cost-benefit side, the system costs less than $10 million a year to operate and it prevents national health expenditures of $291 million a year on average, according to academic research.[36]

The Medicare fraud prevention system helped prevent an estimated $32 million in expenditures in 2012

36 Robert L. Scharff, "A Model of Economic Benefits and Costs from PulseNet," Department of Consumer Sciences, The Ohio State University, Agricultural Research and Development Center, 2010, presentation slide 13, http://bit.ly/18GMGmU.

by provoking changes in medical provider behavior and enabling the Centers for Medicare and Medicaid Services (CMS) to revoke billing privileges, deny claims and suspend payments.[37]

PulseNet also proves that big results don't always require big data. At about 15 megabytes (15 million bytes), its database is dwarfed by some others, for example the Veterans Health Administration's 80-terabyte (80 trillion bytes) corporate data warehouse, which houses information from patients' electronic health records.

The Defense Department (DOD) hasn't yet come up with an estimate of the return on investment for its biometrics program, which cost $3 billion from fiscal years 2007 to 2012.[38] The program collects facial photographs, iris scans and a full set of fingerprints from non-U.S. citizens in Afghanistan, as it did in Iraq until American troops left at the end of 2011.

"Military guys view ROI as some sort of MBA thing that doesn't apply," said John Boyd, director for defense biometrics and forensics, in the office of the Assistant Secretary of Defense for Research and Engineering. "What resonates better, at least within DOD, is more of a risk-assessment standpoint, in other words ... more of an outcome metric."

Without ROI measures, biometrics is at a disadvantage as the Pentagon commences deep post-war budget reductions and weighs programs against one another to determine whether to continue, reduce or stop investment in them. "No objective measurement or understanding of the real or potential value of DOD biometrics exists," a 2012 RAND study found. Consequently, it is "nearly impossible to determine a return on investment—or, more accurately, current results provided by resources expended—beyond anecdotes about 'bad guys' identified."[39]

But without an ROI estimate, the biometrics project is more vulnerable when funding decisions are made. Although DOD says it identified 3,000 enemy combatants among the 1.1 million people from whom biometrics were collected in Afghanistan as of 2012,[40] it's difficult to say whether this is a terrific outcome or merely adequate, or whether it could have been achieved more effectively by other means. This kind of comparison among alternate methods for achieving mission goals is increasingly necessary as budgets shrink. But it's impossible to do without "objective measurement of outcomes and the performance levels of alternative strategies," a 2012 study of federal law enforcement performance measurements found.[41]

Calculating ROI can be especially challenging for programs that use analytics to prevent bad things from happening. "As you're moving toward prevention, one of the important things to measure is costs that are avoided," said Kelly Gent, a leader of CMS's Fraud Prevention System. "If you prevent something from happening, there is nothing to count: That billing did not occur or [those] claims were not denied because they were never made."

That means program staff must come up with novel ways to measure the savings from what wasn't allowed to happen or to continue. "We look at previous billings of providers we removed from the program—we look at their historical billings or those of like providers—and project what would have been paid had we allowed them to remain in the program," Gent said.

Because predictive analytics is just now being adopted more widely in government, methods for estimating its ROI will require continual refinement. For example, the Department of Health and Human Services inspector general questioned CMS's cost avoidance accounting in a review of the 2012 report to Congress on the fraud prevention system; the inspector general found that the $7.3 million CMS reported saving by revoking provider privileges for improper billings might be inflated.

"The department's methodology assumes that not one of the claims submitted by the provider was a legitimate claim," the IG wrote.[42] Yet he found that patients received the same services from other providers after the revocation, proving at least some of the claims were appropriate.

The IG also questioned CMS's claim of saving $68 million by referring suspected fraud to law enforcement agencies. Some cases probably were dropped, he reasoned, while in other cases, fines and penalties for fraudulent activities might have boosted returns higher than accounted for by CMS's estimate.

He also found inaccurate the estimated ROI that CMS reported: $3.30 for every dollar spent on the prevention system. To derive the ROI, CMS divided the total actual and projected savings by a summary of first-year costs, but the agency

37 CMS, Report to Congress Fraud Prevention System First Implementation Year (Washington, DC, Dec. 14, 2012), 24, http://1.usa.gov/1bmPrzB.

38 Shontz et al., *An Assessment of the Assignments and Arrangements of the Executive Agent for DoD Biometrics and Status Report on the Biometrics Enterprise.*

39 Ibid.

40 GAO, *Defense Biometrics: Additional Training for Leaders and More Timely Transmission of Data Could Enhance the Use of Biometrics in Afghanistan.*

41 John Whitley, "Five Methods for Measuring Unobserved Events: A Case Study of Federal Law Enforcement," IBM Center for The Business of Government, October 2012, 10, http://bit.ly/19RgDRu.

42 Health and Human Services Department Office of Inspector General, *The Department of Health and Human Services Has Implemented Predictive Analytics Technologies But Can Improve Its Reporting on Related Savings and Return on Investment* (Washington, DC, A-17-12-53000, September 27, 2012), 6, http://1.usa.gov/1fDyrHa.

overstated the savings in some cases while understating them in others, and failed to report some costs, the IG found.

CMS concurred with the IG and is taking corrective action. Since CMS's fraud protection system is the largest predictive analytics-based program of its kind in government, the IG's recommendations offer practices other predictive analytics users should consider.[43]

For example, he suggested that CMS require its contractors to track the amounts of money recovered as a result of leads generated by the analytics system. CMS also should coordinate with law enforcement agencies to improve the reporting of outcomes of investigations and prosecutions stemming from leads generated by the system. He noted that because CMS used savings—actual and projected—to calculate ROI, it should also have reported all actual and projected program costs.

Preventive analytics programs also might consider adopting ROI estimation methods similar to those proposed by John Whitley in his 2012 study of federal law enforcement agencies. Whitley offers several approaches for gathering and analyzing data to help determine the effect of federal programs on fraud, tax evasion, drug smuggling or other unobserved activities.

He recommends inventorying applicable data available within the agency or from others, and looking for ways to estimate program effects by combining it or enhancing its collection. Surveys and audits can help evaluate how much to invest in different programs, including analytics. Data-driven inspection programs might experiment with secondary screenings to infer their success or failure.[44]

43 Ibid., Appendix: Marilyn Tavenner, administrator, Centers for Medicare and Medicaid Services, letter to Daniel R. Levinson, HHS inspector general.

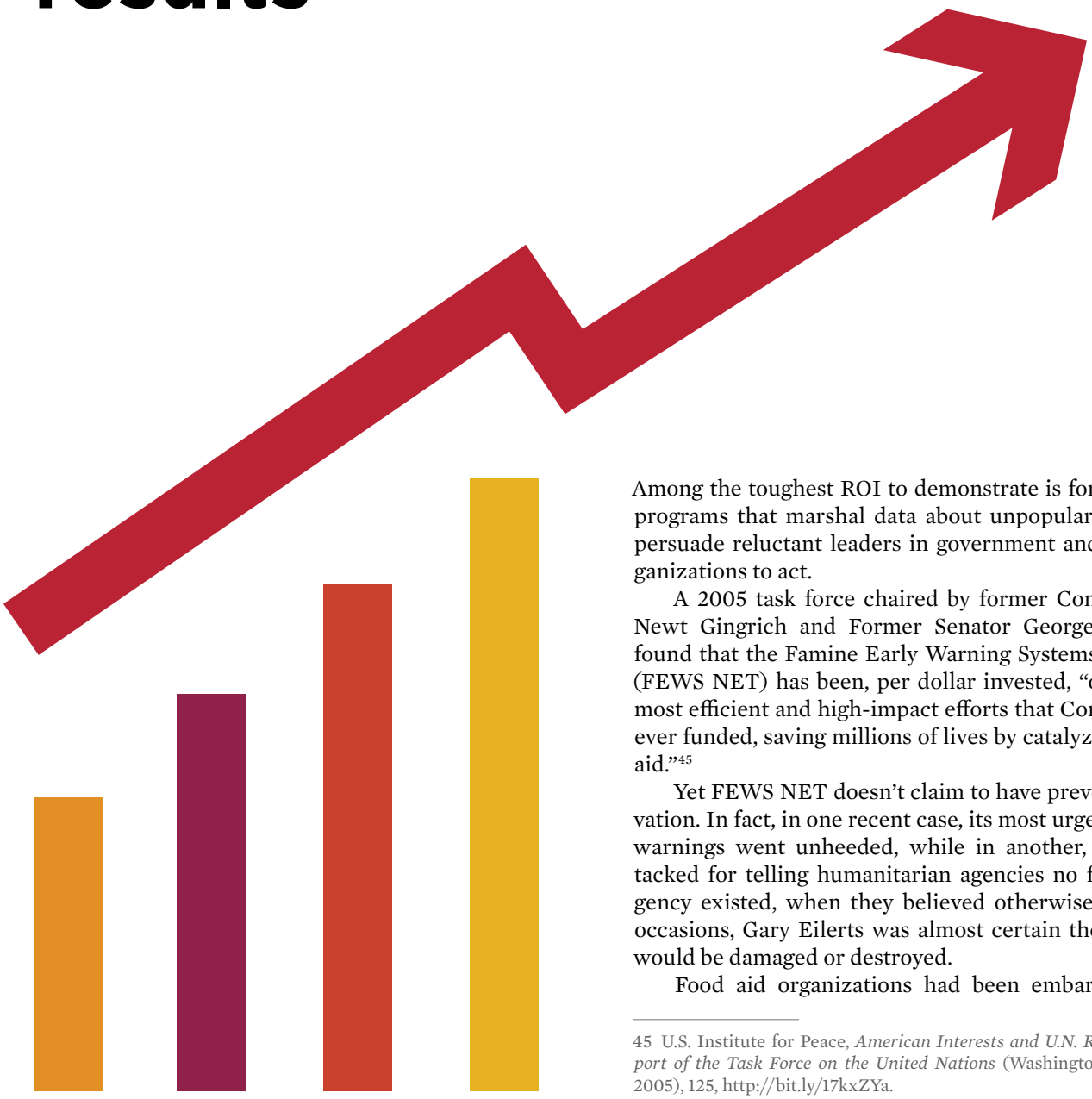
44 Whitley "Five Methods for Measuring Unobserved Events: A Case Study of Federal Law Enforcement," 15-22.

INSIGHTS FOR YOUR ANALYTICS PROGRAM

Long-term data users often had no ROI measures when they began, but developed and adapted them as their projects evolved.

- Most initially reported improved mission outputs and outcomes, for example, the CDC's increasing numbers of foodborne illness outbreaks identified and DOD's numbers of biometrics matched with the subjects on the high-value-target watch list.
- Increasingly, however, they are called upon to deliver cost-benefit and return-on-investment metrics in monetary terms so agency leaders can compare program costs to determine whether data-based efforts are more or less cost effective than alternate strategies.
- To demonstrate ROI, mission analytics programs learned to devote resources to develop data to track financial and other results related in whole or in part to analytics.
- Predictive analytics programs still are refining their cost-benefit metrics and findings and must take care in estimating costs avoided, for example, making certain they report all actual and projected costs.
- To improve their estimates of return on investment, analytics programs can employ surveys and audits, use experimental methods such as secondary screening, and increase and enhance the data they collect.

Give agency leaders clear, concise analysis and proof that analytics are being used to improve mission results



Among the toughest ROI to demonstrate is for analytics programs that marshal data about unpopular truths to persuade reluctant leaders in government and other organizations to act.

A 2005 task force chaired by former Congressman Newt Gingrich and Former Senator George Mitchell found that the Famine Early Warning Systems Network (FEWS NET) has been, per dollar invested, "one of the most efficient and high-impact efforts that Congress has ever funded, saving millions of lives by catalyzing timely aid."[45]

Yet FEWS NET doesn't claim to have prevented starvation. In fact, in one recent case, its most urgent famine warnings went unheeded, while in another, it was attacked for telling humanitarian agencies no food emergency existed, when they believed otherwise. On both occasions, Gary Eilerts was almost certain the network would be damaged or destroyed.

Food aid organizations had been embarrassed by

45 U.S. Institute for Peace, *American Interests and U.N. Reform: A Report of the Task Force on the United Nations* (Washington, DC, June 2005), 125, http://bit.ly/17kxZYa.

their too-slow response to famine in Somalia in 2011. By the time they took notice—after FEWS NET enabled the first-ever real-time declaration of famine in July of that year—tens of thousands of people already had died. The final tally was 258,000 deaths—4.6 percent of the population—due to the food crisis, 10 percent of them children under age five.[46]

FEWS NET had issued 17 increasingly urgent warnings beginning in August 2010, but to many aid officials, dire conditions looked little different than normal for Somalia.[47] What's more, the hardest-hit areas were controlled by al Shabaab, a group the United States had labeled terrorists, and its fighters were killing aid workers. FEWS NET hadn't been able to overcome aid agencies' crisis fatigue and caution.

"When we made the declaration of famine we did have discussions about whether this was going to sink us," Eilerts recalled. "When we tried to convince people, we thought it was a fairly difficult case and people were really not on board. Then, on July 11, when we made the declaration, by that time everybody was starting to see people walking out of Somalia; they were dying on the road. Everybody kind of fell in line."

As John Whitley has pointed out, providing objective data in cases of political contention can "help focus debate on choices between fact-based alternatives."[48] Persistence in presenting data derived from analytics keeps facts front and center and difficult to ignore even when they point to unwelcome conclusions. In Somalia, the famine network was able to perform market analyses that led to delivery of monetary aid for food purchases even when it was almost impossible to send in food itself.

USE DATA TO DRIVE HOME HARD TRUTHS

A few months later, another famine appeared to be brewing in the Sahel—the continent-wide region between the Sahara Desert to the north and the Sudanian savanna to the south—particularly in Niger. This time FEWS NET, almost alone, said the problem was not acute but chronic, and advised against massive humanitarian assistance.

"Everybody was embarrassed by having missed the Somalia issue," Eilerts said. "Then, all of a sudden, within a month or two, the Sahel starts heating up and people say, 'Aha, we're not going to get caught again. We're going to start moving and acting and doing everything we can.' And then here we are saying, 'Wait a minute guys, no, you don't need to.'

"It was just hard for them to rationalize. 'Wait a minute, we just had a famine over here and we missed it, and here's one now and it looks like it's been here forever and it's the same characteristics, and now you don't want us to respond?'"

In the end, it was analytics that saved FEWS NET. The data proved out on the ground.

"We had a lot of data. I went out on the road with a slide show that I showed 50 times. I said, "Here's how this year stacks up against the context," said Eilerts. "What happened was that a lot of the governments out there agreed with us … a lot of the people in the government of Niger even. They said, "You know, this is not a famine.'"

Providing objective data in cases of political contention can "help focus debate on choices between fact-based alternatives."

46 London School of Hygiene and Tropical and the Johns Hopkins University Bloomberg School of Public Health, *Mortality Among Populations of Southern and Central Somalia During 2010-2012* (Rome, Washington, DC, May 2, 2013) 8, http://uni.cf/16DwQNK.

47 Mija-Tesse Ververs, *East Africa Food Security Crisis – An Overview of What We Knew and When Before June 2011* (Assessment Capacities Project, Geneva, 12 July 2011), Annex 1, http://bit.ly/1eUPmDJ.

48 Whitley, "Five Methods for Measuring Unobserved Events: A Case Study of Federal Law Enforcement," 10.

CAPTURE THE BIG PICTURE

Because the famine network doesn't control food aid or development efforts, it can't draw a direct line from its efforts to deaths from starvation prevented. Its power is in presenting data analysis so well and so persistently that it is hard for officials to ignore.

"We've worked really hard in the last year or so on making sure we have four or five sentences that really capture the big picture of what's going on so an undersecretary at the State Department running down the hall or so USAID [officials] can really understand," said Erin Martin, who works with FEWS NET contractor Chemonics.

The company has a Washington, DC-based decision-support team that hones reports from field staff so busy leaders get the message quickly. "They are critical interlocutors in ensuring analysis is solid, clear, substantiated and explained well and clearly," Martin said.

Cara Christie, a USAID consumer of FEWS NET analysis, says the efforts are succeeding. "Having that data and those reports is absolutely critical, not only to planning, but to our confidence about what is going to be needed and what we need to do now to be ready. The level of detail they have on a number of factors that affect food security is rather astounding."

In Somalia, she said, "one of the things we were able to do with FEWS NET was use their information and review the feasibility of using the market and market system to get assistance to the populations that were most in need .... We just peppered them to get really fine analysis of the individual markets. And it worked. I think it's just a tremendous credit to FEWS NET that they were able to provide that in such a strained environment."

Christie views FEWS NET's at first unheeded Somalia warnings as successful. Without them, she says, "I think the fatality figures from Somalia would have been shockingly higher and we probably would not have known what was going on there until significantly later. We would not have known this was a record drought."

The DOD biometrics program hasn't fared as well as FEWS NET in communicating its value. Troops and officers in the field generally support the program but senior officers and civilians don't, according to Boyd. "Training for leaders does not fully support warfighter use of biometrics," the Government Accountability Office (GAO) found.[49]

"My office and others intend to go to the National Defense University, the Eisenhower School to provide seminars to senior leaders on the importance of biometrics," Boyd said. It remains to be seen whether those efforts will be sufficient to win the backing biometrics needs to survive.

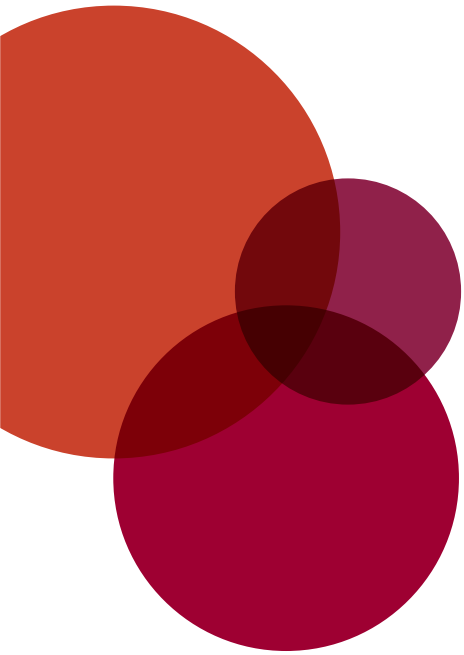
49 GAO, *Defense Biometrics: Additional Training for Leaders and More Timely Transmission of Data Could Enhance the Use of Biometrics in Afghanistan*, 12.

INSIGHTS FOR YOUR ANALYTICS PROGRAM

Data programs with long track records found they had to deliver analysis leaders could use and support.

- The absence of a powerful sponsor can hobble an analytical effort even when it shows mission achievements, as DOD biometrics backers have discovered—especially now, when programs vie for funding as budgets are cut.
- Mature programs struggled when delivering analytics-based messages leaders didn't want to hear, but made headway when they persisted in presenting the supporting data.
- Program managers learned to use leaner, punchier and more visual methods for presenting their findings so senior officials could absorb them and get the main points quickly.
- Programs that grew from the grassroots, such as VHA's VISTA, survived resistance by demonstrating their effectiveness in terms of broad user adoption.

To encourage data use and spark insights, enable employees to easily see, combine and analyze it



In mature data programs, non-managers often applied analytics spontaneously, even before leaders received training on driving data use, so user innovation continues to be welcomed.

"If there are not leaders at the top that want [data-based] information, you're not going to be able to let go of the reins enough to allow the brilliant thinkers at the grassroots to take these tools and provide the types of outputs or analytic products that would change your organization," said Todd Schroeder of the Department of Agriculture's Animal Plant Health Inspection Service (APHIS).

At APHIS, "the idea of analytics took off long before we knew it, in limited locations where larger amounts of data were being collected," he said. "It really started with [business intelligence software]." To save money, the agency introduced the same software across organizations.

So domestic pest surveyors were using it to manage insect traps, while import inspectors were using it to track how many of which types of commodities they had checked and with what results. "People started seeing the linkages," Schroeder said.

They also began asking more of the tool, he said. For example, "[Show me] the imports in a yearly cycle so we can be prepared on what we should be inspecting domestically as a result of possible pests coming into the country, because we're importing more guavas from wherever during the months February and March."

From such demands grew an analytics program that directs inspectors to target the import containers most likely to be carrying pests and lets state and local governments, academics and citizens help catch those that evade inspection by matching them with images on APHIS's online directory.

"We roll a lot of these tools out to the people who are responding to customers on a daily basis. It's the insights and the products they put together to support local needs that provided the insights needed at higher levels to raise new questions," said Schroeder.

User insights also include direct human intervention when no algorithm can do the job. As yet, nothing beats the eye when it comes to verifying "fingerprints," whether of bacterial DNA or human beings.

"Some very bright person said the best way to detect images is the 'eyeball-a-metric' method," said the CDC's Williams. "We've also developed some algorithms to help supplement the eyeball method. Nobody's been able to design one that's actually better than the human eye."

Similarly, the Department of Defense (DOD) uses people rather than formulas to inspect a portion of the fingerprints that must be analyzed. About 9.5 percent

"[Use of analytics changes] daily work because now they're getting more direction on where to go, what to do."

of the fingerprints that came into DOD's Automated Biometrics Identification System (ABIS) from Iraq and Afghanistan through January 2012 couldn't be resolved without being viewed by human forensic examiners, according to the RAND study.[50]

The Centers for Disease Control and Prevention may be able to move humans out of the loop if it adopts whole genome analysis of bacteria instead of today's partial DNA fingerprinting. And if DOD's biometrics program survives, improvements in collection and analytics could mean less analyst intervention is needed. But for now, these and other data-based programs eventually run up against the limitations of the information collected and the capabilities of analytics.

This underscores a point made by Whitley: "All data and analyses are imperfect, contain measurement error and rely on assumptions."[51] That's why analysts "must not overstate the usefulness of their results," and decisions should be informed by analytics combined with "short-run operational realities and constraints, political factors and stakeholder concerns and interests."[52]

KEEP USERS IN MIND

Experienced data users found that employees delivering an agency's core services can be inventive analytics developers, provided the data comes where, when and how they need it to make their work more efficient and effective.

VHA's Stephan Fihn had users in mind when he had his team develop a predictive analytic tool showing which patients are at highest risk of hospitalization, and a data-based scheduling tool for hospital health providers. "We can really assist them in the work of coordinating care, so they're not spending huge time figuring out what data means, or where it is, or tracking it down trying to figure out whether it is relevant or not," Fihn said.

The scheduling tool, known as the patient care assessment system (PCAS), "was specifically built with our [registered nurse] care managers in mind," said Joanne Shear. "It was jointly built with registered nurses from all over the country."

"And nurses are fine-tuning PCAS on the job," said VHA informaticist Tamára Box. "We are constantly assessing how it's being used and the ways that seem to be associated with improvements [in outcomes]."

They keep in mind that forcing nurses and doctors to take extra steps to get to data or analytics can endanger patient safety and create resistance, said Box. "We look at how can we introduce it as a tool that has a benefit to our providers without asking too much of them in return.

"We don't disrupt the workflow," she added.

REDUCE RISK AND VULNERABILITY

Until recently, data at VHA "would trickle down from [a regional] manager after it was provided from the data warehouse in a report through many layers to the people actually using it in day-to-day work," said Fihn. "With trickle-down, people would often feel pressured that the data is more about accountability than work."

Avoiding that kind of pressure is a reason not to impose analytics solely as a performance measurement tool and to make sure employees understand that the goal

50 Shontz et al., *An Assessment of the Assignments and Arrangements of the Executive Agent for DoD Biometrics and Status Report on the Biometrics Enterprise*, 18.

51 Whitley, "Five Methods for Measuring Unobservable Events: A Case Study of Federal Law Enforcement," 30.

52 Ibid.

is improving agency effectiveness, not punishment. "[Use of analytics changes] daily work because now they're getting more direction on where to go, what to do," said APHIS's Schroeder. The risk, he added, is that this gives insights into who is performing well and who isn't, and leaves people feeling vulnerable.

"'If you can tell these [insect trap] surveyors where to go and what to do, what do you need all of these other layers of managers for?' Some people I'm sure are thinking that," he added. He also pointed out that analytics can cause changes in how agencies do business, thereby altering jobs.

"Analytics is disruptive innovation at its finest," Schroeder said. "Maybe it changes the dynamic of how you work .... Maybe you need less inspection because you have more information," he said.

CAPITALIZE ON THE POWER OF THE MISSION

For early analytics adopters, the power of improved delivery of their inspiring missions overcame fear of change. Even usually dispassionate scientists were energized when using analytics meant saving lives.

"Among us are people who appreciate the idea that the work they do makes a difference in programs that touch a large number of people," said U.S. Geological Survey's Verdin of the Famine Early Warning Systems Network. "Some scientists are content to do their research, publish their papers, move on to the next questions. There also are people who like to see their findings make their way into practice and have an impact on decision-making processes. And I'm one of them."

Others were engaged by the opportunity to tackle the world's knottiest problems.

"As a weather guy, I'm very concerned about making sure I provide the best information possible," said Eylander. "I don't have all the answers myself, but they're very, very difficult problems. Weather is such a difficult thing that if you try to do it all yourself you'd just never get there. So the only option you have is to share and collaborate."

"Analytics is disruptive innovation at its finest."

INSIGHTS FOR YOUR ANALYTICS PROGRAM

Projects built on user insights:

- Moved beyond using data exclusively to measure or compare employee and organizational performance by providing tools that enable staff to combine, analyze and use data when, where and how they needed it to speed and ease the work process.
- Were guided by users' insights, implementing good ideas from the grassroots and recognizing those who suggested them.
- Refined analytics tools by watching how employees used them to greatest effect, but without disrupting work flow.
- Made sure those who collected data also benefited directly from it or clearly understood how it improved mission delivery.
- Capitalized on employees' zeal for the agency's mission to help them overcome reluctance about adopting analytics.
- Were honest about the potential for analytics to change agency operations and the jobs of those performing them.

Leaders and managers should demand and use data, and provide employees with targeted on-the-job training

Once early analytics adopters demonstrated the value of data-driven approaches by showing they saved money, improved outcomes or avoided costs, they sought to institutionalize the use of analytics. One sure way to do that was to teach leaders to demand data, they found.

The Veterans Health Administration (VHA) is "grooming a group of leaders who know when to ask for data, know what the analysis should look like," said Gail Graham, who runs VHA's analytics organization.

The effort has produced "leadership who made it very well known that they made decisions based on data," she said. "So if you came in asking for resources, space, people, money, whatever, you were expected to come there with your homework, with the data, and have analyzed the data to support your case."

Long-time analytics users also created centers of excellence devoted to collecting and analyzing data on a large scale and then providing the results and analytics capability organization- and agency-wide. First, however, analytics centers had to learn how the agency operates, and the rest of the organization had to become data literate.

"You have to constantly develop the knowledge base of the infrastructure you're giving people and the people themselves. You have to purposefully grow all of those things," said Graham.

To that end, Fihn's office of analytics and business intelligence at VHA is teaching analytics skills, from the basic, such as making sure each medical center and clinic has a cadre of adept Excel users, to the expert, such as offering university-level courses on advanced analysis. Fihn and his business intelligence team members also offer to speak and present webinars about clinical applications of analytics[53] as part of the VA Information Research Center's extensive online, printed and live educational resources.

At the Centers for Disease Control and Prevention (CDC), analysts have to pass an external course in order to learn how to identify and analyze bacterial DNA data. "It generally takes at least six months of training to develop the expertise to independently classify," said CDC branch chief Peter Gerner-Smidt. They also must pass an external quality check. "It takes a long time to get to this proficiency level," he added.

Failure to properly train employees can undermine both data collection and adoption of analytics, as biometrics leaders at the Defense Department have learned. Because the biometrics initiative is not yet an official program of record, biometrics classes are not offered in Army schoolhouses. As a result, pre-deployment training is ad hoc and troops don't always collect data properly. For example, they have misidentified fingerprints from the left hand as the right and vice versa.[54]

Analytics training enables employees to ask better questions of data, scrutinize it more effectively for patterns and linkages and offers opportunities to improve operations, collect data more efficiently, become comfortable using it and incorporate it in more aspects of their work. Data won't be pervasively used until analytics is standard operating procedure; that can't happen until employees adapt to using data and analytics as part of everyday operations.

CREATE CENTERS OF EXCELLENCE TO SPREAD ADOPTION

Long-standing data programs, such as VHA's, have corporate data warehouses and specialized analytics organizations. Others are creating such centers, some after years of data-driven success.

The Centers for Medicare and Medicaid Services' (CMS) fraud prevention system built in an analytics laboratory from the beginning, said Gent. "We brought in 10 statisticians, economists, programmers who really understand the art of predictive analytics. In program integrity, things evolve and you're forever learning new information and new schemes that need to be identified and targeted."

The Animal Plant Health Inspection Service (APHIS) now is building a business analytics competency center in the hope of helping units across the agency adopt data programs. "The idea is that this information and the use of those analytical products will span all of those functional areas," Schroeder said.

Schroeder also has led efforts to standardize data across APHIS so it can be linked more easily by commodity, geography, type of pest and business entity. "We have to have commonalities in our taxonomy of information to really analyze big sets of data to show us the patterns, themes we might need to be more risk-based, to make better decisions sooner, to have larger impacts," he said.

53 Tamára Box and Stephen Fihn, *Care Assessment Need (CAN) Score and the Patient Care Assessment System (PCAS): Tools for Care Management,* June 27, 2013, http://1.usa.gov/18H54fu.

54 Shontz et al., *An Assessment of the Assignments and Arrangements of the Executive Agent for DoD Biometrics and Status Report on the Biometrics Enterprise,* 34.

INSIGHTS FOR YOUR ANALYTICS PROGRAM

Instilling analytics in all agency activities became a goal once early programs demonstrated gains. It's an ongoing process involving:

- Standardization of data to enable users to look across collections by time, entity, geography, source and other attributes to find linkages and patterns and to share information.
- Formal and on-the-job education.
- Training that's appropriate to the organization and the employee's position; for example, VHA dieticians learn to analyze dietetic data.
- Teaching leaders to base their decisions on data, so they, in turn, require employees to muster analytics to support their cases for funds, staff, space and other resources.
- Centers of excellence with expertise in data analytics, the organization's operations and policies. CMS, for example, houses policy experts along with statisticians in its analytics laboratory. Policy people provide expertise on what is appropriate to bill to Medicare so the fraud prevention system can be trained to identify what isn't.
- Data evangelists who encourage use of data-driven techniques and tools beyond their own units across organizations.

BIGGER DATA, BETTER ANALYTICS

Having experienced the power of analysis when computers were slow, storage limited and databases much smaller, trailblazers now are enhancing and improving their programs as technology evolves. They are embracing bigger data and cutting-edge analytics.

For example, in January 2013, the VHA's Office of Veterans Health Analytics issued a request for information from potential vendors for hardware and software to perform clinical reasoning and prediction. VHA wants the very latest technology and analytical capability, "an emerging class of systems" that use structured data—numbers, dates and strings of words and numbers in a defined format and length—as well as unstructured, such as doctors' notes, text and email.

These systems also must "perceive and adapt to their environment."[55]

The organization wants to enable medical staff to direct the new system using natural language, much as consumers do when they tap questions into online automated assistants on websites. Using regular speech will help more easily identify risks and opportunities to improve patient care—for example, assessing the likelihood a patient may take a fall or be readmitted to the hospital.

DOD's biometrics data is being put to new uses that might ensure the program survives and even expands. In the past several years, defense and intelligence agency analysts have embarked on a new form of sleuthing known as "activity-based intelligence" (ABI). They "approach the data not knowing what they will find," akin to "looking for a needle in a stack of needles to find an unidentified special needle which has some significance," according to an ABI primer in Trajectory, an online magazine.[56]

ABI focuses on patterns of life to identify which activities are normal and which are abnormal and to develop strategy and tactics based on that understanding. It seeks to illuminate relationships between entities—people or vehicles, for example—and their actions. For the most part, ABI involves collecting every kind of intelligence—signals, images, sensor data, human observation, identities, biometrics, forensics—and tagging it to a location.

When the entities are people, the data includes biography—where biometrics comes in—as well as activities and relationships and the environment in which they take place. The key is context, both geographic and relational; the data is limitless.

"The spirit of it breaks the traditional intelligence paradigm," said RAND senior policy analyst Gregory Treverton. "ABI says no, we don't know what we're looking for and by the way, we may find the answer before we know the question. It's not so collection driven … not at all linear."

The increasing availability of data and improved ways of analyzing it also may change how the Centers for Disease Control and Protection (CDC) identifies bacteria causing foodborne illnesses. For example, CDC is exploring next-generation sequencing, a fast, cheap way of analyzing the whole genome of a bacteria, rather than just a partial DNA fingerprint.

"If you look at the sequence of the genome instead, that one is going to be unique for each isolate and you will not be able to make a mistake," said CDC's Gerner-Smidt.

Still, there are challenges to surmount, he said. The current DNA images are only about 10 kilobytes to 25 kilobytes in size, while a single genome sequence takes up megabytes, so CDC still is working on transmission speed and storage. The agency is considering housing genome data elsewhere, perhaps in a public database, such as the National Institutes of Health's National Center for Biotechnology Information. Next-generation sequencing currently is more expensive than PulseNet's current technique, but it can be automated.

And if PulseNet adopts next-generation sequencing, its public-health laboratory partners will need new equipment. "Getting machines in labs to be able to handle the genome will be expensive, but not more expensive than it was outfitting the labs with machines in 1996," said Gerner-Smidt.

So as many federal agencies begin to use data analytics during this era of big data and fast analysis, mature programs are continuing to evolve and adapt. The lessons they learn can help beginners avoid pitfalls, instill analytics faster and move more efficiently and effectively to create data-driven cultures.

55 Department of Interior, RFI-Clinical Reasoning and Prediction Assessment, January 28, 2013, http://1.usa.gov/GzyPGu.

56 Mark Phillips, "A Brief Overview of Activity Based Intelligence and Human Domain Analytics," trajectorymagazine.com, September 28, 2012, http://bit.ly/16R2tNe.

CONCLUSION

The experiences of agencies with mature, data-driven programs reinforce many of the findings of our previous reports: Leaders' attention and support are critical, so make sure the analysis speaks to them; users will make or break the move to data-driven operations, so listen to them, make their work easier and make mission analytics a carrot, not just a stick; find ways to collaborate within and outside your organization to get data, analysis, expertise and even funding.

What early data users didn't do was consciously set out to use "big data." Instead, they asked hard questions and sought data to answer them: How can we detect foodborne illness outbreaks sooner? How can we estimate the quality of a crop months before it is harvested? How can we identify veterans most at risk of hospitalization or death and then target the right care to keep them healthier and at home? How can we focus inspections on containers most likely to hold insects? What patterns of billing and behavior reveal fraud?

Those questions and others propelled these users to collect and analyze data, which then became standard operating procedure and helped their programs evolve.

APPENDIX A
METHODOLOGY

From Data to Decisions III: Lessons from Analytics Pioneers is the third in a series of reports examining federal agencies that have been successful in applying data analytics solutions to further their missions. In the first report, we highlighted promising practices in using analytics to save money, improve services and more effectively achieve agency goals. In the second report, we focused on outlining the steps necessary to begin to use data analytics and how an agency can begin to foster an analytics culture. In this report, we examine how data analytics pioneers began their efforts, how these efforts matured over time, how their ROI is defined and how it helps drive program success.

To accomplish this, we spoke with more than 30 experts from more than 15 agencies, agency subcomponents, offices and private organizations. We conducted more than 50 in-depth phone and in-person interviews between April and August 2013. We attended four local conferences and several online webinars on big data or data analytics. We also conducted an extensive literature review examining recent laws, OMB circulars and memorandums, private sector best practices, news articles and other publicly available agency documentation.

APPENDIX B
FEDERAL INTERVIEWEES AND SURVEY PARTICIPANTS

Association of Public Health Laboratories

Kristy Kubota
Senior Specialist, PulseNet Program

Kirsten Larson
Senior Specialist, Food Safety Program

Cablespeed

Wayne Wheeles
Analytic, Infrastructure and Enrichment Developer, Red Alpha

Centers for Disease Control and Prevention

John Besser
Deputy Chief, Enteric Diseases Laboratory Branch, Division of Foodborne, Waterborne, and Environmental Diseases

Peter Gerner-Smidt
Chief, Enteric Diseases Laboratory Branch

Alvin Shultz
Program Coordinator, Epidemiology and Laboratory Capacity for Infectious Diseases, Division of Preparedness and Emerging Infections, National Center for Emerging and Zoonotic Infectious Diseases

Ian Williams, PhD
Chief, Outbreak Response and Prevention Branch

Chemonics, Overseas Strategic Consulting

Kyle Green
Information and Communications Technology Manager, FEWS NET

Erin Martin
Communications Specialist, FEWS NET

CTOVision

Robert Gourley
Chief Technology Officer, Crucial Point, LLP

DEPARTMENT OF DEFENSE
Department of the Army

John Boyd
Director, Defense Biometrics and Forensics Office of the Assistant Secretary of Defense, Research and Engineering

John Eylander
Cold Regions Research and Engineering Lab, U.S. Army Engineer Research and Development Center

DEPARTMENT OF DEFENSE
Department of the Air Force

Jeff Cetola
Chief, Environmental Characterization Flight
16th Weather Squadron (16WS/WXE), 2nd Weather Group (Air Force Weather Agency)

Chris Franks
Land Information System Team Lead, Air Force Weather Agency

DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Medicare and Medicaid Services

Kelly Gent
Director, Data Analytics and Control Group, Center for Program Integrity

EXECUTIVE OFFICE OF THE PRESIDENT
Office of Science and Technology Policy

George Strawn
Director, National Coordination Office for Networking and Information Technology Research and Development

Wendy Wigen
Technical Coordinator, National coordination Office for Networking and Technology Research and Development

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
Goddard Space Flight Center

Dr. Christa D. Peters-Lidard
Physical Scientist, Hydrological Sciences Laboratory

Office of Management and Budget

Shelley Metzenbaum
Former Associate Director of Performance and Personnel Management
President, The Volcker Alliance

UNITED STATES DEPARTMENT OF AGRICULTURE
Animal and Plant Health Inspection Service

Todd Schroeder
Director, Business Systems Management

Lee Spaulding
Cognos Administrator, Plant Protection and Quarantine

RAND Corporation

Gregory F. Treverton
Director, Center for Global Risk and Security

United States Agency for International Development

Cara Christie
Team Lead, East and Central Africa, Office of U.S. Foreign Disaster Assistance

Gary Eilerts
Program Manager, FEWS NET

United States Geological Survey

James P. Verdin
Physical Scientist, Manager of USGS FEWS NET Activities, Earth Resources Observation and Science

DEPARTMENT OF VETERANS AFFAIRS
Veterans Health Administration

Jack Bates
Director, Business Intelligence Product Line, Office of Information and Technology

Tamára L. Box
Lead, Health IT, Clinical Assessment, Reporting and Tracking Program
National Project Manager, Patient Care Assessment System (PCAS)
Office of Informatics and Analytics, Office of Analytics and Business Intelligence

Stephan Fihn
Director, Analytics and Business Intelligence

Gail Graham
Assistant Deputy Undersecretary for Health for Informatics and Analytics

Joanne M. Shear
Primary Care Clinical Program Manager, Offices of PC Operations (10NC3) and PC Services/Policy (10P4F)

APPENDIX C
CONTRIBUTORS TO THE REPORT

Partnership for Public Service

Anne Laurent, Project Lead

Judy England-Joseph

Cynthia Heckman

Bevin Johnston

Stephanie Mabrey

Seth Melling

Ellen Perlman

Audrey Pfund

Lara Shane

Max Stier

Melissa Wiak

IBM

Daniel J. Chenok
Executive Director, IBM Center for the Business of Government

Gregory Greben
Vice President, Public Sector Business Analytics and Optimization

John Kamensky
Senior Fellow, IBM Center for the Business of Government

Brian D. Murrow
Associate Partner, Business Analytics and Optimization

PARTNERSHIP FOR PUBLIC SERVICE

1100 New York Avenue NW
Suite 200 East
Washington DC 20005

(202) 775-9111
ourpublicservice.org
CFC# 12110

IBM Center for
The Business of Government

600 14th Street NW
Second Floor
Washington DC 20005

(202) 551-9342
businessofgovernment.org