# Roundtable profile:

# Analytics at IDA

**An inside look at the Institute for Defense Analyses and why analytics has always played a vital role.**

**By David E. Hunter**

**Founded in 1956, the Institute for Defense Analyses (IDA)** is a not-for-profit corporation that currently operates three federally funded research and development centers (FFRDCs): the Systems and Analyses Center, the Science and Technology Policy Institute and the Center for Communications and Computing.

FFRDCs are unique, independent entities sponsored and funded by the U.S. government to meet long-term technical needs that cannot be met as effectively by existing governmental or contractor resources. These entities were initially established after World War II as the U.S. government and Department of Defense (DoD) tried to find a way to maintain continued access to the technical and scientific expertise that had proved so valuable during the war effort. IDA's sole business is operating its three FFRDCs.

Collocated with IDA headquarters in Alexandria, Va., IDA's Systems and Analyses Center assists the Office of the Secretary of Defense as well as other government agencies – such as the Department of Homeland Security, the Director of National Intelligence and the Department of Veterans Affairs – in addressing important national security issues, focusing particularly on those requiring scientific and technical expertise. IDA exists to promote national security, preserve the public welfare and advance scientific learning by analyzing, evaluating and reporting on matters of interest to the U.S. government. IDA's goal is to empower the best scientific and strategic minds to research and analyze the most important issues of national security.

To achieve this goal, IDA maintains a highly educated and diverse research staff. In fact, more than 90 percent of IDA's researchers have advanced degrees, with the majority having earned doctoral degrees in a technical field.

Each year, IDA researchers execute hundreds of projects for government sponsors. For each project, research teams comprising the precisely necessary scientific, technical and analytical skills – and with disparate life experiences and backgrounds – are assembled from across IDA's eight research divisions. IDA's flat organization and culture of internal collaboration allow researchers to easily and collegially interact with each other and the Institute's leaders.

Analytics has always played a vital role at IDA. IDA researchers do not use any one specific analytical technique or tool to solve all problems, but rather seek to employ the most appropriate techniques to address each individual research question. Following are some examples of analytical techniques used by IDA researchers to address specific research questions:

*IDA text analytics (ITA).* ITA is a customized software capability, built on proven open source components, for exploratory analysis of highly heterogeneous collections of documents (i.e., exploratory search). It is employed on a wide range of problems at IDA from cybersecurity applications to program evaluation. ITA uses a variety of different techniques based on machine learning and natural language processing to facilitate rapid insight discovery. It supports both search (e.g., looking for specific information) and discovery (e.g., interactive browsing to reveal information for which one may not have even known to look). ITA goes beyond simple keyword search tools through its implementation of analytics-powered facets (or filters), which allow an analyst to view a document set along different dimensions (or through various lenses). These facets, in addition to other visualizations and auto-generated reports, provide rich overviews of the entire information space and can help answer various researchable questions of interest.

ITA utilizes numerous techniques to implement such facets including: key phrase and concept discovery; topic clusters; supervised machine learning facets; customizable entity extractions; and file metadata facets – location, time and format.

ITA is actively developed with new functionality made available regularly such as graph-based visualizations of text corpora, duplicate detection and various other reports to help answer researchable questions.

*Statistical analyses and data mining.* Statistical analyses and data mining are some of the more common analytical techniques employed by IDA researchers. These tools were particularly valuable a few years ago when the Department of Veterans Affairs (VA) asked IDA

# All About the Roundtable

The Roundtable consists of the institutional members of INFORMS with member company representatives typically the overall leader of O.R. activity. The Roundtable is composed of about 50 organizations that have demonstrated leadership in the application of O.R. and advanced analytics. The Roundtable culture is peer-to-peer, encouraging networking and sharing lessons learned among members.

The Roundtable meets three times a year. Roundtable goals are to improve member organizations' OR/MS practice, help Roundtable representatives grow professionally and help the OR/MS profession to thrive. Further information is available at http://roundtable.informs.org.

The Roundtable also has an advisory responsibility to INFORMS. According to its bylaws, "The Roundtable shall regularly share with INFORMS leadership and advise the INFORMS Board on its views, its suggested initiatives and its implementation plans on the important problems and opportunities facing operations research and the management sciences as a profession and on the ways in which INFORMS can deal proactively with those problems and opportunities." The Roundtable meets with the INFORMS president-elect each spring to discuss practice-related topics of interest to him or her, and with the entire INFORMS Board each fall to discuss topics of mutual concern.

This series of articles aims to share with the INFORMS membership at large some information and insights into how O.R. is carried on in practice today.

to investigate the causes of perceived inequities in the VA's disability compensation program. This multi-billion dollar program provides monthly payments to military veterans with injuries or disabilities incurred or aggravated during military service.

The IDA research team met with VA leadership, traveled across the country to interview hundreds of claims adjudicators, and – perhaps most importantly – collected and analyzed data on millions of disability compensation awards. During this project, IDA researchers formulated hypotheses based on their gained understanding of the VA adjudication process. Further, they employed advanced data mining and exploratory data analysis techniques to find additional factors and interactions implicit in the data.

From the hypotheses and data, IDA employed statistical analyses to test each hypothesis and to quantify the amount of the observed variations that is accounted for by each factor. The IDA analysis identified the main factors contributing to the observed variation, dispelled some common misperceptions, and made policy recommendations to further improve the equity and consistency of disability compensation awards.

*Econometrics and optimization.* Of the roughly $500 billion dollar annual defense budget, about 20 percent is used for procurement – buying systems for our armed forces to use, as opposed to developing new systems or maintaining the systems we already have. What we buy ranges from bullets to ballistic missiles, half-ton trucks to M1 tanks, and inflatable rafts to aircraft carriers. About $50 billion per year is spent on major defense acquisition programs (MDAPs), the most sophisticated and most expensive military systems. These are the nation's investment portfolio against future military operational needs. Understanding the cost, schedule and risks of any one major program is complicated. Understanding their interactions and behavior as a portfolio is even more daunting.

More than **90 percent** of IDA's **researchers** have **advanced degrees,** with the majority having earned **doctoral degrees** in a **technical** field.

## FFRDCs: Unique capabilities

"FFRDCs were established to provide the Department of Defense with unique analytical, engineering and research capabilities in many areas where the government cannot attract and retain personnel in sufficient depth and numbers. They also operate in the public interest free from organizational conflicts of interest and can therefore assist us in ways that our industry contractors cannot."

**Hon. Ashton B. Carter**
Under Secretary of Defense
Acquisition, Technology and Logistics

**IDA modeled** the **cost and benefits** of early **warning** and **detection** technologies employed **to defeat** biological weapons **attacks on** major **U.S. cities.**

IDA has been working with DoD to develop and improve a decision-support tool that models the cost and schedule of all MDAPs simultaneously. This tool, called "PortOpt," allows DoD analysts to predict the likely cost and schedule impact of proposed procurement schedule changes, and to find practical schedules that minimize total procurement cost across all programs, given a fixed budget and fielding requirements. It also provides a means to estimate the overall cost and schedule impact on existing programs of adding a new program or cancelling a program. These capabilities have direct applications to affordability analysis, portfolio analysis, and reprogramming in response to unexpected budget reductions. PortOpt gives DoD the ability to identify – in days or weeks, rather than weeks or months – opportunities for savings, feasible responses to disruptions or impending budget crunches.

At the heart of PortOpt are two key analytical tools. The first is an econometric model of how future procurement costs for each program would vary as a function of production schedule. Because this is a causal model, sophisticated statistical techniques are required to distinguish the effect of schedule on cost from the equally common effect of cost on schedule, or the effect of technical challenges on both. The second key tool is a large mixed integer linear program (MILP) that approximately describes the problem of finding the minimum-cost set of simultaneous schedules subject to constraints on annual budget, latest permitted fielding dates, minimum and maximum production rates, plant capacity constraints and practical limits on which production schedules could be implemented in real life. The MILP uses piecewise-linear approximations to the econometric cost functions, resulting in a formulation with thousands of binary variables, tens of thousands of continuous variables, and tens of thousands of constraints.

*Discrete event simulation modeling.* IDA uses discrete event simulation modeling to assess defense weapon systems. A suite of similar models called IMEASURE, built with ExtendSim, examines various aircraft types; IDA has used these simulation models to examine fighters, helicopters, cargo aircraft and unmanned aerial systems.

Given a particular system's reliability and maintainability (RAM) performance and a target operational capability metric, the model can be used to independently estimate maintenance manpower requirements by job specialty and/or appropriate spare stock levels. Alternatively, given a particular set of available maintenance manpower and spares stock, the model can assess the system's operational capability (mission capable rate, sortie generation rate, operational availability). We often use the model at IDA to make assessments and predictions of operational test performance or to estimate program unknowns (e.g., manning or sparing resource requirements) to support independent cost estimates. Aside from the RAM inputs, there are many additional data and modeling assumptions required to run the model (aircraft turn durations, abort rates, mission schedule, etc.), contributing to the intractability of solving such analytical problems without simulation.

*Cost/benefit analyses.* For the U.S. Department of Homeland Security, IDA modeled the cost and benefits of early warning and detection technologies employed to defeat biological weapons attacks on major U.S. cities. This work involved modeling the dispersion of aerosolized pathogens in various venues such as an outdoor park in Chicago, O'Hare Airport and Grand Central Terminal in New York City. The lifecycle costs and benefits (i.e., reduced mortality and morbidity) of these technologies were simulated over a range of pathogens, venues and operation cycles. The results suggest that net present value of all the technologies was positive.

### The Future

Over the past 60 years, IDA researchers have been asked to provide independent analytic assessments and analyses on a wide range of public policy questions. In fact, the variety of interesting work is one of the oft-mentioned reasons why so many talented people enjoy working at IDA. The future looks to be no different. While it is not possible to predict with certainty the specific research questions that IDA will be asked, they will most assuredly continue to involve some of the more critical aspects of national security. And IDA analysts and researchers will continue to leverage the latest analytic techniques to provide government decision-makers with high-quality independent assessments. **OR/MS**

*David E. Hunter* (dhunter@ida.org) is an assistant director in the Cost Analysis and Research Division (CARD) at the Institute for Defense Analyses (IDA), as well as IDA's representative to the INFORMS Roundtable.