Megatrends paper 2014-02

"Big Data"

Background:

"Big data" refers to sophisticated data-parsing tools and enormous datasets. How big is big? Data from social-media traffic, mobile transactions, and GPS coordinates alone generate over 2.5 quintillion bytes daily.[1] Moreover, the amount of digital information is doubling every two years, and experts predict total volume increasing fifty times by 2020.[2]

The process to transform vast volumes of raw data into decision-quality information comprises four steps. First the data is digitized and structured. Next, the data is "scrubbed" to remove errors, placed into standard forms and tagged with metadata. Then the scrubbed and tagged data is made available to analysts (e.g., via networks). Lastly, analysts apply algorithms to the data.[3]

An algorithm is a sequence of steps and instructions that parses data to identify patterns and relationships.[4] Anyone who has shopped on Amazon.com has encountered the output of an algorithm—in this case, called "item-to-item collaborative filtering." The algorithm matches each customer-purchased item with similar items, and then combines those similar items into a recommendation list. Amazon's algorithm stands out among other e-commerce filters because of its scalability over very large customer bases (tens of millions) and product catalogs (millions), sub-second processing times, and real-time updating capability, which combine to make consistently high-quality recommendations, as judged by customers' feedback. [5]

Government agencies are also committing significant resources to leverage big-data analytics. The Department of Defense's (DOD) Advanced Research Projects Agency (DARPA) is developing advanced analytic and search capabilities for intelligence analysts. Testifying in March 2014 before a House Subcommittee on Intelligence, Dr. Arati Prabhakar, Director of DARPA, noted his agency "seeks to enable analysts to make sense of the huge volumes of intelligence-rich information available to them from existing sensors and data sources." Included in this effort is DARPA's Memory and Exploration of the Internet for Defense, (MEMEX) that seeks to develop an Internet search capability to find information in the "deep Web"—the parts of the Web not indexed by standard commercial search engines. MEMEX's initial focus will be human trafficking, which is a factor in many types of military, law-enforcement, and intelligence investigations, and which has a significant Web presence to attract customers.[6]

Big data and the analytics available to unlock its secrets have attracted opponents who are concerned about legal, security, and privacy ramifications. For example, the Electronic Privacy Information Center notes that "while there are many benefits to the growth of big data analytics, many traditional methods of privacy protections fail. Many notions of privacy rely on informed consent for the disclosure and use of an individual's private data. However, big data means that data is a resource that can be used and reused, often in ways that were inconceivable at the time the data was collected."[7]

Effect on the Information Environment (IE):

DOD is spending millions on these initiatives because of the likely impact big data will have on the future IE. The IE is the realm which encompasses where and how information is collected and analyzed, and C2 is exercised. In future conflicts, victory may go to the force that can leverage the exponential increase in data volume at a pace at which events unfold.

In recent years, the volume of data collected within this dimension has far outpaced the ability to analyze and disseminate it in actionable form. For example, in the year of the World Trade Towers attack, the State Department processed over seven million visas and over 24 million overseas travelers landing at US airports.[8] [9] The challenge of protecting the homeland today is even greater. Visa issuances in 2013 topped nine million.[10] Every day, two million passengers fly into, within, or over the United States.[11] Department of Homeland Security officials note that the challenge is not simply to find the "needle in the haystack." "Protecting the homeland often depends on finding the most critical needles across many haystacks—a classic big data problem."[12]

Compounding the problem of volume is the increasing rapidity with which events unfold. This in turn often requires responses measured in minutes if not seconds. A study conducted under the auspices of the Royal United Services Institute (RUSI) adds that it is a problem that will not go away. RUSI notes, "the expectations of commanders continue to grow as more digital sensors and collectors enter service and as new signatures are identified that need to be 'washed' against multiple databases. These expectations will not be met without corresponding improvements in tools and techniques to support the search and analysis of data." The study goes on to recommend conducting proofs of concept for leveraging big data in support of military operations. [13]

Recent Examples:

The past decade plus of war saw modest advances in the joint force's use of big data in operations within the IE. Human terrain teams in Afghanistan leveraged computing power to import traditional databases, biometric data, and imagery to extract information to improve ISAF's understanding of the local population and key relationships within it.[14] By fusing human-terrain information with coordinates of previous engagements, data from topographical and street maps and phone data, units were able to predict within 700 meters where insurgents' supplies for making roadside bombs were likely to be stored. [15]

Big data research continues today at NATO's Allied Rapid Reaction Corps HQ to better identify patterns in data which may be useful for predicting likely targets for radicalization. The effort aims to, in part, substitute human analytical function with computing power to speed the intelligence cycle process. [16]

Big data is also being put to work to enhance computer security. Vendors such as McAfee, Inc., continue to improve capabilities for analyzing huge volumes of security data and identify anomalies from baselined normal activity that could signify the presence of malware. In addition to detecting external breaches, vendors are also making strides in identifying insider threats. The technology company Sphere of Influence developed a product from a scientist's premise that "insiders," such as NSA contractor Edward Snowden, behave differently than their co-workers, and, from such behavior, subtle anomalies can be detected. The technology monitors network

activity, collects users' behaviors, and builds sophisticated models of an entire organization. From these models, an advanced machine learning layer detects insider threats and alerts officials for the need to investigate.[17]

Implications for Future Joint IO Force:

As technologist Paul Ford notes, "data that's well defined and cleanly organized can be connected to whole other swaths of data. So once you build big organized indexes of human beings—one of their search terms, one of their phone calls, for example— you can merge them into one mega-index. And you can combine that mega-index with other mega-indexes. There's enormous power in linking things."[18]

It is these linkages that make the value of big data more valuable than the data itself. "Big data technologies can derive value from large datasets in ways that were previously impossible; indeed, big data can generate insights that researchers didn't even think to seek."[19] The idea is rooted in what the intelligence community calls "mosaic theory," whose origins predate by decades the advent of ubiquitous computing and large data sets. The theory recognizes the value of aggregated, seemingly innocuous, data which reveals over time insights greater than any individual piece of information can.

The mosaic theory in the era of big data raises privacy issues which will have significant implications for the future joint force. Indeed, as recent unauthorized disclosures about NSA data gathering has revealed, legal and policy rulings are already having an impact on defense and intelligence agencies procedures for using big data. Looking further ahead the RUSI report suggests that, as armed forces conduct "war amongst the people," additional legal issues pertaining to how information is collected, analyzed, and stored will surface.[20]

It is in today's scientific research and e-commerce arenas where one can begin to envision the operational and privacy implications of big data for the joint IO force. Algorithm-enhanced mosaic theory was on display during a recent research event sponsored by Nokia in which researchers filtered GPS and cell phone tower data. From a user's location history alone, the researchers were able to estimate the user's gender, marital status, occupation, and age. Moreover, the researchers were able to predict a user's likely *future* location by observing past location history. [21]

Mega-indexes and algorithms could also be a boon for Military Information Support Operations (MISO) planners in conducting target-audience analysis. Such tools can help answer questions such as what the apparent goals, motivations, and characteristics of groups in an operating area are, and how susceptible are they to influence activities.[22]

Answers to these questions typically involve segmenting a large population within an operational area into smaller sub-populations based on multiple variables such as age, educational levels, ethnicity, etc. The value of continued advances in big data analytics could be twofold: "finer grain" segmentation, and a corresponding reduction in time required to accomplish segmentation.[23]

Additionally, the availability of vast amounts of geo-locational data combined with analytic software will provide IO planners the ability to tailor the content and timing of influence

campaigns more than ever before. Already today, technologies can determine with pinpoint accuracy personal location even within buildings where GPS signals are notoriously weak. Shopkick, for example, is a mobile phone application that allows retail merchants to track their customers from the moment they walk in via in-store devices that pick up humanly inaudible sounds emitted by the customer's mobile phone microphone. Customers subsequently are rewarded for their visit by receiving store coupons good for future redemption.[24] Might the future joint IO force offer rewards to mobile phone owners who provide data on the location of IEDs or locations of terrorists?

Once themes and messages are disseminated, big-data analytics can enhance the assessment of influence operations effects. Just as businesses today use sentiment analysis from social media to measure the response to marketing campaigns and adjust course accordingly, IO force of 2020 may be able to do so as well—and with unmatched speed.

McKinsey cites the continuing development and refinement of a variety of tools for providing real-time monitoring and response to Web-based consumer behavior and choices.[25] Media Intelligence, for example, uses such tools to provide real-time monitoring and evaluation of print, online, broadcast, and social media mentioning a company's brand. To protect the good name of a brand, it tracks more than 50,000 online news outlets, 3,000,000 social-media sources, and 1,000 broadcast and print outlets in more than 80 countries using natural-language processes to analyze the tone of references to the company.[26]

Big data's implications on the joint IO force are not limited to target-audience analysis, messaging, and assessment, however. Its capability also has implications on resource allocation and command and control. Technologist Kenneth Cupier provides a glimpse into the power of big data in resource allocation in a business setting. He chronicled a leading chemical company whose sales volumes in their existing segmented territories had plateaued. Using big data, the company was able to dice its existing seven territories into 70 "micro-markets." It then pulled reps away from over-served markets and redeployed them to underserved markets. Within a year, the sales-growth rate doubled—without an increase in marketing or sales costs. Just like the commercial sector, such technology-enabling resource allocation decisions will benefit global and theater-level force managers.[27]

Big data also offers the possibility of achieving a vision first articulated by Admiral Cebrowski over 15 years ago—self synchronization. He defined the term as "the ability of a well-informed force to organize and synchronize complex warfare activities from the bottom up."[28] While trust between a commander and subordinates is a prerequisite for implementing a bottom-up approach to warfighting, so is a an enhanced understanding of the information environment, which a combination of a common operational picture (COP) that the joint force's next generation C2 system—the joint information environment—and big data analytics can provide.

Technologist Dan James provides a scenario that involves a form of self-synchronization called swarming. The concept involves the use of autonomous or semi-autonomous individuals or groups to conduct simultaneous actions from multiple directions to accomplish an objective. As the term suggests, the concept evokes an image of how insects go about nest building or food gathering. In James' scenario, US monitoring of social media in a volatile country detects an increase in GPS-enabled mobile devices indicating an impending threat to a diplomatic post.

Alerted to this development, a JFC launches UAVs and cues other sensors to determine the presence and intent of threatening crowds. This data is then combined with other databases that include the location and capabilities of host-nation military and police units. The resulting big-data-enabled predictive analysis of crowd movements enables the host nation to quickly deploy its forces from all directions to cordon off the diplomatic post and either prevent larger crowds from forming or disperse those that already have.[29]

Major David Faggard (USAF), however, outlines an alternative scenario based on a real-world event that doesn't play out for a future JFC as well as James'. Faggard envisions an "e-citizen" force armed with the full computing power of mobile technology with real-time network updates. Such a well-organized, technically savvy group conducting social swarming within the physical and cyberspace domains could overwhelm and paralyze their opponents' command and control and decision making.[30]

He goes on to cite a real-world incident that provides a glimpse of the future he predicts. In 2008, ten terrorists attacked a hotel in Mumbai, India; over 160 people died. While far from overwhelming Indian authorities' decision making, the terrorists were able to keep pace with India's situational awareness. Their Pakistan-based commanders continuously monitored broadcast news and Internet reporting, and provided situation updates to them via BlackBerrys, satellite phones, and GSM handsets.

The more the future joint force relies on big data, the more attention must be given to ensuring terrorists and adversary nations cannot glean detailed knowledge into the workings of the tools and techniques for leveraging big data. For example, there are several different algorithms capable of performing a data sorting function. If an adversary were to determine how the joint force sorts its big data, it may be possible to predict the likely output of a query which, in turn, could increase the probability that the adversary may correctly forecast a JFC's course of action. Similarly, as potential adversaries increasingly rely on big-data analytics, US and allied knowledge of their algorithms, coupled with the use of Red Teams, might yield valuable insights into an adversary's next moves.[31]

DOTMLPF Change Recommendations:

Doctrine (to include TTP):

- D.1. Develop and promulgate TTP and include in joint and allied doctrine the lessons learned by human terrain teams, NATO's Allied Rapid Reaction Corps, and others in the joint force having first-hand experience using big data during the past decade of war.
- D.2. Work with allies to conduct proof-of-concept demonstrations of big-data techniques which may yield further TTP and doctrinal insights. Include in the proof of concept an examination of big-data policy and legal implications.

Training:

- T.1. As part of a proof-of-concept demonstration, assess training needs necessary for the intelligence and operations communities to leverage big data.

- T.2. Encourage combatant commands and component organizations to design exercises and other training events in which an adversary's employment of social media and other readily available technologies challenges the blue force's ability to maintain situational awareness, and command and control forces.

Materiel:

- M.1. Obtain situational awareness and monitor DARPA and other DOD labs' research into big data. As appropriate, recommend programs and further areas of investment in DOD's IO investment strategies, and provide advocacy throughout the planning, programming, and budgeting process.
- M.2. Conduct a quick look of the capabilities of two or three commercially available tools for conducting one or more of the following IO and IRC areas: computer security, sentiment analysis, geo-locational data.

Functional capability: Data Gathering and Information Processing-Sense-making

Applicable JCAs: Collection; Analysis; Organize Information; Develop and Share Knowledge and Situational Awareness; Monitor

---

[1] For an extensive, non-technical treatment of the subject see McKinsey Global Institute (MGI), "*Big Data: The Next Frontier for Innovation, Competition and Productivity*", May 2011, white paper.
[2] Prediction on volume by 2020 was made in 2011; see IDC, 'Digital Universe Study: Extracting Value from Chaos', June 2011.
[3] Douglas W. Hubbard, *How to measure anything: Finding the value of intangibles in business* (New York: Wiley, 2010).
[4] Executive Office of the President of the US, *Big Data: Seizing Opportunities, Preserving Values*, May 2014
[5] Greg Linden, Brent Smith, and Jeremy York, "Amazon.com Recommendations", IEEE Internet Computing, Jan-Feb 2003.
[6] Dr. Arati Prabhakar, Prepared Testimony for Subcommittee on Intelligence, Emerging Threats & Capabilities, U.S. House of Representatives, March 26, 2014.
[7] See http://epic.org/privacy/big-data/#overview for additional information on privacy implications of big data
[8] Visa data taken from 2001 State Department Report of Immigrant and Nonimmigrant Visas issued at Foreign Service posts for Fiscal Years 1997-2001.
[9] Figures provided by Department of Commerce, Travel and Tourism Section. Includes data only from participating airlines.
[10] Visa data taken from State Department Report of Immigrant and Nonimmigrant Visas issued at Foreign Service posts for Fiscal Years 2009-2013.
[11] Executive Office of the President of the US, *Big Data: Seizing Opportunities, Preserving Values*, May 2014
[12] Ibid.
[13] Neil Couch and Bill Robins, *"Big Data for Defence & Security";* prepared for Royal United Services Institute (RUSI), September, 2013.
[14] RUSI.
[15] Economist Magazine, The Science of Civil War, April 21, 2012.
[16] RUSI.
[17] SOI news release; "Sphere of Influence says Insider Threats are Detectable," *Business Wire*, June 18, 2013.
[18] Paul Ford, "Balancing Security and Liberty in the Age of Big Data," *Bloomberg Business Week*, June 13, 2013.
[19] Executive Office of the President of the US, *Big Data: Seizing Opportunities, Preserving Values*, May 2014.
[20] War amongst the people is a term coined by General Sir Rupert Smith. It connotes that future joint force engagements will increasingly take place among, against, and in defense of civilians. Civilians will be the targets and the objectives to be won, as much as the defeat of an opposing armed force or capture of territory. See JIOWC

megatrends paper on ubiquitous computing and Smith's and Dr. Ilana Bet-El's *Military Capabilities for War Amongst the People*, e-publication, accessed 27 February 2014 at www.strategycenter.org.
[21]Steven Bellovin, Tony Jebara & Sebastain Zimmeck, *When Enough is Enough: Location Tracking, Mosaic Theory, and Machine Learning* (working paper, 2013).
[22] JP 3.13.2, *MIS.*
[23] MGI.
[24] MGI.
[25] MGI.
[26] See: http://www.nasdaqomx.com/corporatesolutions/public-relations-solutions/monitoring/media-intelligence-old.
[27] Kenneth Cukier, "The Rise of Big Data," *Foreign Affairs*, May/June 2013.
[28] See B.J.A van Bezooijen; unpublished paper, "Military Self-synchronization: An Exploration of the Concept"; delivered at 2007 Naval Postgraduate School C2 Symposium.
[29] Stew Magnuson, "Defense, Intel Communities Wrestle With the Promise And Problems of 'Big Data'," *National Defense Magazine*, March 2013.
[30] Major David Faggard, (USAF)," Social Swarming: Asymmetric Effects on Public Discourse in Future Conflict," *Military Review*, March-April 2013
[31] While dated, the Defense Science Board Task Force report on *The Role and Status of DOD Red Teaming Activities*, September, 2003, provides an overview of the types and purposes of red teams.