CHAPTER 16

Future Trends

"Progress has not followed a straight ascending line, but a spiral with rhythms of progress and retrogression, of evolution and dissolution."

Johann Wolfgang von Goethe

"[I]t is pretty easy to go far astray when trying to project the future of technology. In the cases of data mining and predictive analytics, however, the future is becoming reality at such a rapid pace that almost anything that I write will be outdated before the first copy of this book is purchased. Therefore, I will confine my comments to a few areas that I am particularly excited about, even if that "future" represents current reality. In many ways, that is one of the features that make this area of research and practice so exciting."[1]

Just as I "predicted" in the first edition of this text, several of the emerging technologies and "future trends" have become mainstream capabilities and the same cautions apply now. Text mining, fusion centers, and data warehouses – all relatively new capabilities a few years ago – have been extended and in some cases even replaced with tools like sentiment analysis, Big Data, and cloud computing. The following sections include an overview of new or emerging capabilities that I believe are particularly interesting and/or promising. This list is not meant to be all inclusive or comprehensive, particularly given the high probability that some of the most promising capabilities that we are likely to incorporate into the applied public safety and security setting over the next few years are still sitting on the work benches of data science "imagineers" as I write this chapter.

16.1 [REALLY] BIG DATA

Really massive data sources have been around for many years. Particularly as we look outside of our own domain, we can see other professionals, including those in marketing and financial services, who have been working to effectively manage and exploit very large datasets. Similarly, scientists working in astrophysics and high-energy physics have been developing more effective methods of weak signal detection in extremely large data; capabilities with direct applicability and benefit to us.

While Big Data really are not new, the capabilities and associated tradecraft developed to effectively exploit these resources are relatively new and increasingly accessible to the data science community. Moreover, the formalization of concepts associated with big data, particularly as relates to the three "Vs" (volume, velocity, and variety)[2] can help us better manage and use this resource. Finally, considering "big data" as a unique space also offers new opportunities for transdisciplinary collaboration to identify potentially useful data science capabilities from other disciplines facing similar challenges in the management and meaningful exploitation of "big" data that can be applied to operational public safety and security analysis.

Looking ahead, the increased deployment of sensors will markedly grow both the amount and associated granularity of data in support of even deeper analysis. In particular, the increase in wearable devices like Google Glass and other persistent collection tools, including geoenabled devices like smartphones will contribute to a "people as sensors" model. This will generate massive amounts of location intelligence, including streaming video and other content that can be merged and integrated in support of activity-based intelligence (ABI) and pattern of life analysis. Extending this concept, Google cofounder and CEO Larry Page has envisioned the seamless integration of Google in the human brain, so that "[w]hen you think about something…you will automatically get information,"[3] a model of "supreme artificial intelligence" that approaches the cyborgs of science fiction. While these sources are promising as relates to our understanding of the real complexity of behavior, the data collected will be truly massive, and the associated management and exploitation requirements will be staggering.

16.1.1 Biometrics

Biometrics to include fingerprints and DNA have become accepted standard practice in the operational public safety and security domain. DNA databases in particular represent a great source for the identification, as well as exclusion of criminal suspects.[4] Ongoing development of additional biometric capabilities include face and voice recognition, as well as iris scans.

As technology has developed, biometrics are increasingly easy to collect and use for identification purposes. Many of us are familiar with, and may even have direct experience with the use of fingerprints for access control at secure facilities. Other benefits and uses, however, quickly come to mind. For example, the perpetration of identity theft would be significantly curtailed, because while I might be able to steal your credit card number, or even your personally identifiable information in an effort to assume your identity, I cannot easily steal and use your fingerprints, iris scan or genetic profile. In response, insurance companies and financial services institutions increasingly are moving to the use of biometrics for identification purposes.[5]

As we consider the expanded use of biometrics, however, several additional opportunities and related questions emerge. For example, will there be a day when we totally abandon credit and ATM cards, and use biometric data exclusively for this purpose? Will saying "Colleen McCue" into a voice recognition device at the point of sale be sufficient to make a purchase? While it would ostensibly make theft more difficult, criminals could still perpetrate fraud in other ways. It also may be increasingly difficult to establish and/or assume an identity or legend in support of undercover or covert operations if biometric data and technology become prevalent.

Again, these uses of biometric data could be used to create some tremendously granular data in support of activity-based intelligence; however, what privacy considerations would need to be considered regarding access to and use of these data? Data privacy already is being actively debated as relates to the use of automated license plate recognition (ALPR) and toll collectors. However, these data sources are relatively crude and unreliable for establishing actual location for a specific individual. On the other hand, while I can loan you my car, I cannot loan you my fingerprint or DNA. Biometrics would represent a significant step in the ability to accurately identify and locate an individual. Similar to earlier debates regarding the routine collection and use of DNA for law enforcement purposes, these are all issues that are likely to be debated before any widespread public safety and security adoption and use.

16.2 ANALYSIS

Big data is not so much about big data as it is about an enhanced ability to extend and more effectively realize the promise of predictive analytics. Therefore, concomitant developments in analytics to include new tradecraft, technology, and even better approaches to processing will enable us to realize the promise of big data.

As we consider future trends and capabilities in analysis, it is important to revisit basic concepts regarding data mining and predictive analysis. While exciting new methods, technology, tradecraft, and even nomenclature (e.g., "data science") have been developed, predictive analytics at its core still goes back to confirmation and discovery; *confirmation*, operationalization, and extension of what we know or think we know, and *discovery* of new trends, patterns, and relationships. Within this basic framework of characterization, confirmation and discovery, some analytic methods and capabilities merit additional attention given their promise for enhanced approaches to crime and intelligence analysis.

16.2.1 Geospatial

At the time of the first edition of this text, I had some familiarity with the use of GPS data to identify location. This knowledge was based on a specific case, and at the time that it was used to identify the location of a stalker it was still relatively unusual in local law enforcement, and represented a protected, law enforcement sensitive capability. Since this time, the use of GPS and other location-based capabilities to find people, including missing persons, represents a resource that is discussed openly and described in detail in the media. Underscoring the growth and development of geospatial capabilities, as well as the marked increase in the distribution and use of geoenabled mobile phones, it is not unusual for the public to call early and often for the analysis of mobile phone data to triangulate location in an abduction or other missing persons case; the same "sensitive" capability that was used just a few years earlier in the stalking case.

16.2.2 Activity-Based Intelligence

Building on the geospatial capabilities introduced and used throughout this text, activity-based intelligence is an emerging methodology introduced in response to the need to effectively leverage GEOINT into actionable intelligence, ultimately yielding "big value" from "big data"[6] by effectively capturing and modeling pattern of life. By using advanced analytics to "identify patterns, trends, networks, and relationships hidden within large, data collections from multiple sources: full motion video, multispectral imagery, infrared, radar, foundation data, as well as SIGINT, HUMINT and MASINT information,"[7] we can better anticipate and influence future behavior, particularly within a geospatial context.

16.2.3 Experts, Expert Systems, and the Power of the Crowd

If data mining and predictive analytics truly are game changing, why have they not been universally adopted? It would seem that increased public safety is something that everyone could get behind; however, there has been a lag in the acceptance of automated tools in some areas. Research from the political science community may provide an answer to this apparent disconnect between science and practice. It seems that people are more inclined to trust an "expert" despite the finding that the accuracy of "expert" predictions does not differ from those of mere mortals, both of which perform well below predictions derived using statistics and mathematical modeling.[8]

I have direct personal experience with this phenomenon. Several years ago, I attended a scientific meeting that included a lively debate over expert opinion versus statistical estimates of risk for future violence. Despite the fact that the data overwhelmingly supported the accuracy and reliability of the statistical estimates, the attendees found a number of exceptions that would have been missed by computer models and ultimately elected to stay with the human judgments. One possible explanation for this is that people may find comfort in the authority that an "expert" conveys, rather than believing that human nature can be reduced to math and equations.[9] Given the capacity that data mining and predictive analysis can bring to support public safety and security, however, this disconnect between science and practice really needs to be addressed. Perhaps the best model for the paradigm shift required lies somewhere in between those two extreme positions and could include *domain* experts using the expert systems embodied in data mining and predictive analysis software.

16.2.3.1 Consensus Opinions

Although the Defense Advanced Research Projects Agency (DARPA) FutureMAP program was cancelled due to public outrage over government-sponsored wagering on future terrorist attacks and assassinations, consensus opinions have been used with some success. In a unique application of Bayes' theorem, naval scientist John Craven used consensus expert opinions to locate the US nuclear submarine *Scorpion.*[10] Bayesian inference is particularly appealing for applied public safety and security analysis because it supports the incorporation of tacit knowledge and domain expertise from experts representing diverse backgrounds, potentially bringing the "best of all worlds" to the analytical process.

16.2.3.2 Crowd Sourcing

Again, Ushahidi, TomNod, and even astronomy efforts like the Andromeda Project[11] and Galaxy Zoo,[12] are leveraging the power of the crowd to address big data processing tasks. Similarly, the DARPA Red Balloon Challenge effectively used crowd sourcing to solve a "distributed, time-critical, geolocation problem"[13] with implications for public safety and security, as well as search and rescue. Not only can these capabilities be used to break up tasks, but as TomNod has demonstrated, there are certain tasks that humans are uniquely suited to perform. Using so-called, "artificial intelligence" TomNod has been able to effectively leverage the crowd in support of search and rescue, feature extraction, and a number of other challenging geospatial tasks.

16.3 OTHER USES

Again, looking outside the specific public safety and security domain, innovation in other professional disciplines frequently can be leveraged to support crime and intelligence analysis techniques. For example, with access to incredibly complex data resources, Google has developed a business model that enables them to infer age, gender, and interest based on online behavior in support of increasingly targeted advertisements.[14] In keeping with the concept of ABI or pattern of life analysis, location intelligence also is being increasingly incorporated into these models.[15] The ability to effectively characterize the "when, where, and what" of consumer behavior supports optimization, including microtargeting; something that is also being used in political campaigns[16] as staff develop approaches to anticipation and influence. All of these capabilities have direct and obvious implications for crime and intelligence analysis

Unfortunately, like many other tools, these capabilities also can be used against us. For example, if our adversaries can obtain sufficient data either through direct collection or theft (e.g., stolen marketing databases), can they also use analytics to infer identity, preferences, and/or pattern of life in support of increasingly complex attacks? Will ABI be used to support, augment, or even replace hostile surveillance? While the credit card companies are able to generate new credit cards and provide credit monitoring services in response to data theft, it is not quite clear how we will address the theft or compromise of more complex data sources, particularly those relating to pattern of life.

16.4 TECHNOLOGY AND TOOLS

In addition to data and analysis, new technology and tools, as well as more efficient analytics service delivery models will greatly increase both the capacity and capability of crime and intelligence analysis.

16.4.1 Domain-Specific Tools

The emergence of tools designed specifically for public safety and security analysis continues to parallel general advancements in data science. While many of the examples included in this book were generated using technology and tradecraft that were developed originally for some other domain, the development of applications designed specifically for public safety and security analysis makes them even more accessible to the crime and intelligence analyst. Similarly, advances in the visual depiction of complex analytical output continue to be the focus of research and development. Analytical output that directly addresses the "I'll know it when I see it" metric and build on the end user's tacit knowledge in support of direct transfer to and use in the applied setting continue to represent a powerful trend in the industry.

16.4.2 Processing

New capabilities like the SAS High Performance analytics and SAP HANA leverage in-memory processing in support of real-time or near-real-time analysis of the really big data that we are increasingly encountering – effectively enabling "analysis at the speed of thought."[17] As illustrated by the "Food Truck" example in Chapter 14, the ability to provide real-time feedback regarding suspicious activity will significantly inform and enhance surveillance detection operations, while also providing meaningful insight and situational awareness directly to forward deployed operational personnel. Similarly, the use of SAS HP analytics brings the promise of real-time analysis of transactional data in support of truly effective approaches to fraud detection and prevention, including real-time scoring of transactional data, thereby eliminating many of the serious inefficiencies associated with the traditional "pay and chase" model.

16.4.3 IBM Watson

Since its public debut on Jeopardy, Watson has evolved from a very exciting science project into something that is being used to solve real problems in a variety of different and disparate domains. Of particular interest to the operational public safety and security community is Watson's agility with transactional data, as well as its powerful capabilities to infer associations and relationships. By using Semantic Analysis Technology (SAT), Watson is able to identify subtle and nuanced associations in data, including patterns of behavior suggestive of financial crimes and fraud.[18] Similarly, Watson is able to leverage life-event detection and psycholinguistic tools in support of more comprehensive and nuanced pattern of life analysis, including inference regarding which social media accounts might be yours.[19] Again, the applicability of these capabilities to operational crime and intelligence analysis is exciting.

16.4.4 Managed Service, Software as a Service, and Cloud Computing

Managed service delivery models, Software as a Service (SaaS), and cloud computing minimize, if not completely eliminate the "burden of ownership" associated with purchasing, installing, and maintaining sophisticated analytic software locally. Moreover, leverage of the analytic fusion center model also offers the promise of optimized analytic resources, including skilled personnel, as well as the benefits associated with data and operations that are both vertically and horizontally integrated.[20]

16.5 POTENTIAL CHALLENGES AND CONSTRAINTS

In addition to new data sources, technology, and tradecraft, a number of potential challenges and constraints have emerged including transnational crime, and concerns regarding the protection of privacy, civil rights, and civil liberties. In addition, the rapid proliferation of analytic capabilities also creates new challenges for the crime and intelligence analyst as they work to acquire technical proficiency with specific sources and methods, while remaining open to the "art of the possible" as embodied in new data, technology, and tradecraft. Again, any potential challenge also represents a unique opportunity to succeed and I look forward to the development of meaningful solutions that will further enhance, inform, and strengthen crime and intelligence analysis going forward.

16.5.1 Globalization of Crime

The increasingly global nature of our world has created numerous opportunities as relates to fluid, continuous pathways for commerce, speedy response to manmade and natural disasters, and a larger sense of community that transcends national boundaries. Unfortunately, this increased interconnectedness has concomitantly cultivated and grown the transnational nature of crime, which will represent a significant challenge going forward. We know that criminals in the United States frequently exploit jurisdictional boundaries and differential enforcement as a means by which to enable their criminal activity, particularly organized crime. Similarly, national boundaries, differential enforcement, and even differences in language, culture, and the rule of law create an attractive environment that can be exploited by criminals. Further complicating this challenge is the role that international crime, terrorism, and violent extremism play in the creation of ungoverned and under-governed spaces that support or otherwise enable crime by limiting the ability to effectively and consistently enforce the rule of law.[21]

Again, location matters and the important role that location plays in crime has been highlighted by example throughout this text. Transnational crime not only exacerbates but also perpetuates the challenge of ungoverned or under-governed space by creating veritable "enforcement-free zones" for all manner of crime including illegal supply chains that move drugs, guns, people, bulk cash, and other natural resources, and the violence used to enforce the rules, norms, and boundaries associated with these locations, further threatening vulnerable populations and fragile states. Solutions will not be simple or easy; likely requiring an unprecedented level of global collaboration and cooperation in support of meaningful and long lasting approaches.[22]

16.5.2 Let the Process Guide the Solution

> "If all you have is a hammer, everything looks like a nail."
>
> **Abraham Maslow**

The rapid proliferation of crime and intelligence analysis tools has been both a blessing and a curse. A blessing in their ability to effectively translate high-quality advanced analytics for use in the operational public safety and security environment. A curse because they have created a growing population of crime and intelligence analysis "technicians," or individuals with deep expertise regarding a specific source or method but limited knowledge regarding analysis as a process and the importance of context in the interpretation and effective use of the results. While this may provide immediate, short-term benefit to the organization, it also threatens to limit the ability to effectively respond to new or emerging threats, as well as the ability to successfully adopt new sources and/or methods as they become available.

It is not unusual for an analyst to become exceptionally proficient in a specific technology, method, tool, or tradecraft to the exclusion of others. In this situation, however, if all I have is the equivalent of an analytic hammer, then every question begins to assume the form of a nail (or is forced to fit that model). This may occur out of comfort and complacency on the part of the analyst. On the other hand, organizational commitment to a specific tool or platform, particularly if it is perceived as being "cutting edge" and/or was expensive, can similarly constrain access. Unfortunately, this situation does not enable the analyst to effectively adapt and respond to the exceptionally diverse and evolving nature of our problem space. The better approach would be to train the equivalent of analytic master carpenters who would be able to select the tool most appropriate for the question. Ideally, these analysts also would be uniquely suited to quickly embrace and use new tools as they become available, and may even be involved in the development of next-generation tools; either through transdisciplinary adoption and use of existing capabilities from other domains, or even *de novo* development of novel capabilities that would benefit our community.

As discussed in Chapter 4, "wicked" problems also create a unique hazard for the analyst in that the favored solution tends to drive the definition and characterization the problem, which may result in analysis confounded by "circular logic."[23] Moreover, there are no perfect solutions, no "free lunches" in analysis.[24] Ultimately, it is important to remember that these tools support the data mining *process*, and while they might be necessary for analysis, they certainly are not sufficient for the insight required to support meaningful and effective anticipation and influence in the operational public safety and security environment. Therefore, the primary objective for most technology solutions is that it will optimize human time and attention by surfacing interesting patterns, trends, and relationships. Again, it is the domain expertise and ability to create operationally relevant and actionable output that is the priceless element in the applied public safety and security analytical process. As noted in the opening to this chapter, new data sources, technology, and tradecraft are being developed daily. The well-trained analyst will be able to seamlessly adopt these new capabilities as they are made available, as well as those that have not even been invented yet; assuming a fluid approach to technology and allowing the questions to guide the analytic approach, thereby letting the problem guide the solution.

16.5.3 Privacy and Civil Liberties

The challenge of applying eighteenth century legal statues and concepts to twenty-first century technology has created tension for policy makers and analysts alike. Unfortunately, common misperceptions regarding data mining and predictive analytics, and related concerns regarding privacy and civil liberties, as well as uncertainty regarding the overall value of this approach in operational public safety and security threaten to severely curtail the use of advanced analytics for crime and intelligence analysis.

Other new capabilities have encountered resistance when they were first introduced, including the use of DNA evidence and criminal investigative analysis, or the behavioral analysis of violent crime. In these cases, the community was well served by individuals and professional organizations that were not only willing to go out and advocate, but also document successes in support of these capabilities. These groups established standard practice and related credentialing requirements for expert testimony, while concomitantly educating the legal community, juries, their peers, and the public. Underscoring the success of these efforts, juries have come to expect the introduction of DNA evidence at trial, regardless of the crime being tried, as a result of the so-called "CSI Effect." Perhaps we will see a similar embrace on the part of operational security analytics.

Therefore, as we increasingly use data mining and predictive analysis to anticipate, predict, and prevent crime, we must be sensitive to the concerns regarding the protection of privacy and civil liberties. Some of the specific issues are described here.

16.5.3.1 Data

In response to several high-profile data breaches, as well as revelations regarding government collection and use of data, the public is becoming increasingly savvy about their data. Areas of new or increased concern include personally identifiable information (PII), sensitive financial information, social media,[25] and communications "metadata," as well as location data.[26] On the other hand, individuals increasingly view their data from a transactional perspective, expressing a willingness in both word and action to exchange their information for something of value.[27] In other words, they are willing to give up some privacy in exchange for a benefit. Whether it is a financial incentive or information regarding restaurants within their vicinity based on geolocation data, they frequently are willing to barter their data in exchange for desired information, goods, or services.

Automatic license plate recognition and automated toll collection systems provide a good example of the complexity of these issues, particularly as relates to automated collection capabilities. Even recently, it was not unusual for police departments and other agencies to collect the license plate information in an effort to identify specific vehicles associated with a particular event or location. While this was not incontrovertible evidence, it was possible to use this information to make some inferences regarding specific individuals. For example, there is no expectation of privacy associated with a vehicle in a public location and it was not unusual to collect license plate numbers from organized crime or gang funerals in an effort to identify putative membership and better understand these networks. ALPR systems and other sensors, however, can automate this process, significantly increasing the amount of data collected, as well as concomitantly increasing the speed and relative complexity of the associated analysis.

Similarly, is there a difference between seeing an individual in a particular public location or using GPS-enabled capabilities like those embedded in a smartphone or other device to infer location? And does this matter if it is a vendor or other commercial entity that is collecting this information for market analysis purposes, or a law enforcement or security organization supporting public-safety-related analysis or an active investigation? In response to these concerns, there has been increasing public debate, as well as proposed legislation[28] that will curtail or otherwise restrict access to information. Again, law enforcement agencies had been collecting these data for years; however, the introduction of automated systems has prompted discussion of what is appropriate and what may require additional legal review and permission, particularly as relates to persistent collection.

16.5.3.2 Analysis

Recently, data scientists in a number of disparate domains are learning painful lessons regarding the power of advanced analytics and that just because you can do something does not necessarily mean that you should. While it appeared that most attention in previous years was directed at public safety and security use of these tools, more recently marketing and retail use of predictive analytics have received considerable negative attention. This has been particularly true of derived products and the use of advanced analytics for decision support. Perhaps one of the most high profile examples came from Target,[29] a retailer that developed an algorithm to infer the early stages of pregnancy based on purchasing behavior. They then used this information to create personalized packages of coupons and related product offerings in an effort to gain market share in the lucrative pregnancy and childcare market. This practice came to light after the retailer mailed a brochure with pregnancy and baby-related coupons to a high school student, which prompted an uncomfortable discussion in her home and subsequent public outrage over Target's use of advanced analytics when the story came to light.

This example has surfaced a number of issues regarding the legal, ethical, and even practical boundaries related the use of predictive analytics as the public increasingly views some of the results as inappropriate transgressions of privacy. Many of the comments from the data science community suggested that there is nothing inherently "bad" about what Target did, noting that it was just math. Rather, it was how the results were presented and used that created the issue. Again, all of the data used in the Target analysis had been collected legally; however, the subsequent analysis created a derived product that was significantly more sensitive than the original sources. Moreover, analysis, even if based entirely on open source data, that reaches into private family matters including marriage and family planning crosses important social boundaries; transgressions that may have the unintended consequence of repelling possible customers rather than expanding market share. Therefore, at least one of the lessons learned in response to the Target case has been that just because you can do something does not necessarily mean that you should, and related but more subtle considerations regarding operational use, and associated messaging of analytic results and context.

We have a similar challenge in operational security analytics that emerged in 2003 when the Data Mining Moratorium Act was introduced, which proposed that the use of advanced analytics in the applied public safety and national security domains should be significantly regulated or even curtailed. As with the marketing example, the relevant data sources and analytic techniques were legally and ethically available. Rather, it was the use of the results that created concern. Therefore, the related take home message for the crime and intelligence analyst is to think about what you are doing – particularly as relates to the creation of derived products. These are sensitive for a number of reasons. Second, these are increasingly important capabilities. Misuse, whether intentional or otherwise, creates challenges for all of us and may limit our ability to use some of the most powerful tools available to us if they are not used properly, and with sensitivity and respect.

16.5.3.3 "Prediction" and the Minority Report Concern

Additional public concern relates to the use of data mining and predictive analytics to guide action, particularly as it relates to taking action based on statistical "predictions" or what *might* happen versus addressing actions already committed (e.g., the "Minority Report" concern).[30] Most of the examples outlined in this text rely on the use of analysis to identify general patterns and locations of future risk based on behavior; however, data mining and predictive analytics also are being used to develop models designed to identify specific individuals who may merit additional scrutiny (e.g., behavioral screening at airports), as well as individual risk assessment models (e.g., recidivism). Again, people were being investigated, questioned, and detained based on what they might do, sometimes wrongly, rather than what they have done before data mining and predictive analytics came into use.[31] This issue is not particularly new or unique to data mining and predictive analytics. Rather, it goes back to responsible decision making in the operational law enforcement and public safety community; something that should be respected regardless of the source of the opinion or related analysis.

16.5.4 First Do No Harm…

> "Rather than thinking about exceptional moral rules for exceptional moral situations we should almost always see exceptional moral situations as opportunities for us to show exceptionally-deep commitment to our deepest moral values."[32]

Again, the unique circumstances of crisis and conflict mapping merit special consideration. As has been noted throughout the text, adding location not only increases the potential value of the data but also may concomitantly increase the sensitivity of the data and related derived products. In crisis and conflict mapping in particular, this "value add" may pose harm to already vulnerable populations. This is an issue that the community is working through currently;[33] however, as discussed with the Target example, just because you can do something does not always mean that you should. The analyst should always maintain awareness of the situation and larger context associated with their efforts, and consider the broader implications and use of their work.

While these very effective tools hold great promise, they may not be available to us for long given concerns about privacy and civil liberties. It is our responsibility, therefore, to use these tools ethically and responsibly, adhering to the spirit as well as the letter of the law. Adoption and use of advanced analytics in the operational public safety and security setting represents a major paradigm shift for the community. It is important that we clearly recognize what these tools can and cannot accomplish, though. There are no crystal balls, no "Minority Report." At their foundation, this is all just math. These tools are a means to an end, with the major objective being enhanced public safety and security. With that in mind, we need to understand and respect the issues and related concerns. In addition, we need to educate ourselves and the operational end users, as well as the public regarding the value and limits of these tools, and the important protections in place, to ensure truly informed debate and appropriate use within the applied setting. Going forward, please be good stewards of these capabilities and make an effort to create "informed consumers" among your clients, end users, and the general public, and remember the following:

- Reporting, collecting, and compiling data are necessary, but not sufficient to increasing public safety.
- Advanced analytics are used in almost every segment of society to improve service delivery and optimize resources.

- Operational security analytics support the meaningful exploitation of public safety and security data necessary to information-based anticipation and influence, including prevention, response, and consequence management.
- Used responsibly, operational security analytics can enhance public safety, prevent crime and change outcomes.

IS DATA MINING EVIL?

Further confounding the question of whether to acquire data mining technology is the heated debate regarding not only its value in the public safety community but also whether data mining reflects an ethical, or even legal, approach to the analysis of crime and intelligence data. The discipline of data mining came under fire in the Data Mining Moratorium Act of 2003.

Unfortunately, much of the debate that followed has been based on misinformation and a lack of knowledge regarding these very important tools. Like many of the devices used in public safety, data mining and predictive analytics can confer great benefit and enhanced public safety through their judicious deployment and use. Similarly, these same assets also can be misused or employed for unethical or illegal purposes.

One of the harshest criticisms has addressed important privacy issues. It has been suggested that data mining tools threaten to invade the privacy of unknowing citizens and unfairly target them for invasive investigative procedures that are associated with a high risk of false allegations and unethical labeling of certain groups. The concern regarding an individual's right to privacy versus the need to enhance public safety represents a long-standing tension within the law enforcement and intelligence communities that is not unique to data mining. In fact, this concern is misplaced in many ways because data mining in and of itself has a limited ability, if any, to compromise privacy. Privacy is maintained through restricting access to data and information. Data mining and predictive analytics merely analyze the data that are made available; they may be extremely powerful tools, but they are tools nonetheless. With data mining, ensuring privacy should be no different than with any other technique or analytical approach.

Unfortunately, many of these fears were based on a misunderstanding of the Total Information Awareness system (TIA, later changed to the Terrorism Information Awareness system), which promised to combine and integrate wide-ranging data and information systems from both the public and private sectors in an effort to identify possible terrorists. Originally developed by DARPA, this program was ultimately dismantled, due at least in part to the public outcry and concern regarding potential abuses of private information. Subsequent review of the program, however, determined that its main shortcoming was related the failure to conduct a privacy impact study in an effort to ensure the maintenance of individual privacy; this is something that organizations considering these approaches should include in their deployment strategies and use of data-mining tools.

On the other hand, some have suggested that incorporation of data mining and predictive analytics might result in a waste of resources. This underscores a lack of information regarding these analytical tools. Blindly deploying resources based on gut feelings, public pressure, historical precedent, or some other vague notion of crime prevention represents a true waste of resources. One of the greatest potential strengths of data mining is that it gives public safety organizations the ability to allocate increasingly scarce law enforcement and intelligence resources in a more efficient manner while accommodating a concomitant explosion in the available information –

the so-called "volume challenge" that has been cited repeatedly during investigations into law enforcement and intelligence failures associated with 9/11. Data mining and predictive analytics give law enforcement and intelligence professionals the ability to put more evidence-based input into operational decisions and the deployment of scarce resources, thereby limiting the potential waste of resources in a way not available previously.

Regarding the suggestion that data mining has been associated with false leads and law enforcement mistakes, it is important to note that these errors happen already, without data mining. This is why there are so many checks and balances in the system – to protect the innocent. We do not need data mining or technology to make errors; we have been able to do that without the assistance of technology for many years. There is no reason to believe that these same checks and balances would not continue to protect the innocent were data mining to be used extensively. On the other hand, basing our activities on real evidence can only increase the likelihood that we will correctly identify the bad guys while helping to protect the innocent by casting a more targeted net. Like the difference between a shotgun and a laser-sited 9 mm, there is always the possibility of an error, but there is much less collateral damage with the more accurate weapon.

Again, the real issue in the debate comes back to privacy concerns. People do not like law enforcement knowing their business, which is a very reasonable concern, particularly when viewed in light of past abuses. Unfortunately, this attitude confuses process with input issues and places the blame on the tool rather than on the data resources tapped. Data mining can only be used on the data that are made available to it. Data mining is not a vast repository designed to maintain extensive files containing both public and private records on each and every American, as has been suggested by some. It is an analytical tool. If people are concerned about privacy issues, then they should focus on the availability of and access to sensitive data resources, not the analytical tools. Banning an analytical tool because of fear that it will be misused is similar to banning pocket calculators because some people use them to cheat on their taxes.

As with any powerful weapon used in the war on terrorism, the war on drugs, or the war on crime, safety starts with informed public safety consumers and well-trained personnel. As is emphasized throughout this text, domain expertise frequently is the most important component of a well-informed, professional program of data mining and predictive analytics. As such, it should be seen as an essential responsibility of each agency to ensure active participation on the part of those in the know; those professionals from within each organization that know where the data came from and how it will be used.

Unfortunately, serious misinformation regarding this very important tool might limit or somehow curtail its future use when we most need it in our fight against terrorism. As such, it is incumbent upon each organization to ensure absolute integrity and an informed decision-making process regarding the use of these tools and their output in an effort to ensure their ongoing availability and access for public safety applications.

16.6 CLOSING THOUGHTS

> "Information analysis is the brain of homeland security. Used well, it can guide strategic, timely moves throughout our country and around the world. Done poorly, even armies of guards and analysts will be useless." Markle Foundation's Task Force on National Security in the Information Age[34]

This statement continues to be true and emphasizes the critical importance of sound crime and intelligence analysis. While new technology, tools, and tradecraft can enhance our ability to effectively ask and answer the "hard questions," we still need to do more than collect data. As analysts, we need to analyze it in a way that yields the meaningful insight that will translate directly into information-based decisions and operational support.

In closing, "[p]redictive analytics…does not provide guarantees. Instead, it is all about increasing the likelihood that a desired outcome will occur – at the right time, the first time. These concepts of increased likelihood and timeliness are what make applying it to decision making so enticing."[35] These are very worthy, yet attainable goals for operational public safety and security analysis. With that objective in mind, I wish you well, and encourage you to go forward and do good.

Bibliography

1 McCue C. Data mining and predictive analysis: intelligence gathering and crime analysis. Burlington, MA: Butterworth-Heinemann (Elsevier); 2007. p. 315.

2 Dumbill E. What is big data? O'Reilly Radar. http://strata.oreilly.com/2012/01/what-is-big-data.html; 2012 [accessed 11.01.2012]

3 Farber D. At 15, Google's ambitions remain unbridled. CNET, September 27. http://www.cnet.com/news/at-15-googles-ambitions-remain-unbridled/; 2013.

4 McCue C, Smith GL, Diehl RL, Dabbs DF, McDonough JJ, Ferrara PB. Why DNA databases should include all felons. Police Chief 2001; 68: 94–100.

5 Collier K. TELCOs and other businesses are preparing to store their biometric data such as voice fingerprint records to identify customers. Herald Sun, March 24. http://m.heraldsun.com.au/news/victoria/shops-and-telcos-collecting-fingerprints-voice-records-of-customers/story-fni0fit3-1226863673700?sv=56d3cd522b852c4520ade640b93eba71&nk=9f770a49e74bac80d5ca091371d1f78d; 2014.

6 Long L. Remarks as prepared, Letitia A. Long, Director, National Geospatial-Intelligence Agency, SPIE 2013 Defense, Security + Sensing Symposium, May 01. https://www1.nga.mil/MEDIAROOM/SPEECHESREMARKS/Pages/SPIEDSSSymposium.aspx; 2013.

7 Long L. Remarks as prepared, Letitia A. Long, Director, National Geospatial-Intelligence Agency, SPIE 2013 Defense, Security + Sensing Symposium, May 01. https://www1.nga.mil/MEDIAROOM/SPEECHESREMARKS/Pages/SPIEDSSSymposium.aspx; 2013.

8 Colvin G. (2006). Ditch the 'experts.' Fortune 2006; February 6, p. 44.

9 Ibid.

10 Sontag S, Drew C, Drew A. Blind man's bluff: the untold story of American submarine espionage. New York: HarperCollins; 1999.

11 Kelley P. Crowdsourcing the cosmos: astronomers welcome all to identify star clusters in Andromeda galaxy. University of Washington, December 04. http://www.washington.edu/news/2012/12/04/crowdsourcing-the-cosmos-astronomers-welcome-all-to-identify-star-clusters-in-andromeda-galaxy/; 2012.

12 Adams T. Galaxy Zoo and the new dawn of citizen science. The Guardian, March 17. http://www.theguardian.com/science/2012/mar/18/galaxy-zoo-crowdsourcing-citizen-scientists; 2012.

13 Tang JC, Cebrian M, Giacobe NA, Kim H-W, Wickert D. Reflecting on the DARPA Red Balloon Challenge. Commun ACM 2011; 54(4): 78–85.

14 An interesting parlor game, you can see your Google profile and better understand how it sets the ads that you see. In my case, Google has determined my age correctly, believes that I am a male, and have an interest in astronomy (Read M. How old does Google think you are? Gawker.com, January 27. http://gawker.com/5879895/how-old-does-google-think-you-are; 2010; Stanley C. Find out how old Google thinks you are (among other things). Flavorwire, January 27. http://flavorwire.com/253616/find-out-how-old-google-thinks-you-are/; 2012).

15 Brustein J. If your phone knows which aisle you're in, will it have deals on groceries? Bloomberg Businessweek, January 06. http://www.businessweek.com/articles/2014-01-06/apples-ibeacon-helps-marketer-beam-ads-to-grocery-shoppers-phones; 2014.

16 Issenberg S. How President Obama's campaign used big data to rally individual voters. MIT Technol Rev. http://www.technologyreview.com/featuredstory/509026/how-obamas-team-used-big-data-to-rally-voters/; 2012 [accessed 19.12.2012].

17 Few S. Data analysis at the speed of thought. InformationWeek, 01 March, http://www.informationweek.com/software/information-management/data-analysis-at-the-speed-of-thought/d/d-id/1030748?page_number=2; 2005.

18 Cohan P. How you can profit from Watson's 'Jeopardy' win, Daily Finance, 30 March. http://www.dailyfinance.com/2011/03/30/how-you-can-profit-from-watsons-jeopardy-win/; 2011.

19 Brownlee J. IBM's next big thing: psychic Twitter bots. Fast Company, March 03. http://www.fastcodesign.com/3025738/ibms-next-big-thing-psychic-twitter-bots; 2014.

20 McCue C, Miller L, Lambert S. The Northern Virginia military shooting series: operational validation of geospatial predictive analytics. Police Chief, February. http://www.policechiefmagazine.org/magazine/index.cfm?fuseaction=display&article_id=2871&issue_id=22013; 2013.

21 Miles D. Ham: Africa presents opportunity, challenges. U.S. Department of Defense. June 19. http://www.defense.gov/news/newsarticle.aspx?id=116802; 2012.

22 Miklaucic M, Brewer J. Convergence: Illicit networds and national security in the age of globalization. Washington, DC: National Defense University Press; 2013.

23 Rittel H., Webber M. Dilemmas in a General Theory of Planning, Policy Sciences 1973, 4, p. 166.

24 Braun ML. Data analysis: the hard parts. Marginally Interesting: Machine Learning, Computer Science, Jazz, and All That. http://blog.mikiobraun.de/2014/02/data-analysis-hard-parts.html; 2014 [accessed 17.02.2014].

25 Omand D, Bartlett J, Miller C. "A balance between security and privacy online must be struck…." DEMOS. http://www.demos.co.uk/files/_Intelligence_-_web.pdf?1335197327; 2012.

26 The Centre for Spatial Law and Policy, legal and policy issues associated with geospatial data and technology (http://spatiallaw.com).

27 Morris T. 3 ways to personalize the customer experience without getting to personal, parature, June 10, http://www.parature.com/personalize-cx/; 2013; Cisco. Cisco customer experience research: automotive industry global data. http://www.cisco.com/web/about/ac79/docs/ccer_report_manufacturing.pdf; 2013.

28 The Geolocation Privacy and Surveillance (GPS) Act, and related geolocation privacy legislation (http://www.gps.gov/policy/legislation/gps-act/).

29 Duhigg C. How companies learn your secrets. The New York Times. http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=all&_r=0; 2012 [accessed 16.02.2012]; also, the talk that triggered the controversy: Pole A. How Target gets the most out of its guest data. Predictive Analytics World, http://rmportal.performedia.com/node/1373; 2010.

30 Stroud M. The minority report: Chicago's new police computer predicts crimes, but is it racist? The Verge, February 19. http://www.theverge.com/2014/2/19/5419854/the-minority-report-this-computer-predicts-crime-but-is-it-racist; 2014.

31 For review of the issue, McKinney JM. Washington State's return to indeterminate sentencing for sex offenses: correcting past sentencing mistakes and preventing future harm. Seattle Law Rev 2002; 26: 309–336, http://digitalcommons.law.seattleu.edu/cgi/viewcontent.cgi?article=1747&context=sulr

32 Bioethicist Dr. Lachlan Forrow, director of ethics and palliative pare programs, Beth Israel Deaconess Medical Center – in e-mail to Sheri Fink (12NOV90), as quoted in: Fink S. Five days at memorial. New York: Crown; 2013. p. 468.

33 DETECTER, Detection Technologies, Terrorism, Ethics, and Human Rights (http://www.detecter.eu)

34 The Markle Task Force on National Security in the Information Age, including James B. Steinberg, Vice President and Director, Foreign Policy Studies. Protecting America's freedom in the information age 2002, Markle Foundation.

35 Bernstein D. Big data's greatest power: predictive analytics. SAP, 22 November. http://blogs.sap.com/innovation/big-data/big-datas-greatest-power-predictive-analytics-01138403; 2013.