MITRE | Center for Strategic Competition

OCCASIONAL PAPERS, VOL. 1, NO. 6 - NOVEMBER 10, 2022

USING PUBLICLY AVAILABLE INFORMATION IN AMERICAN "WHOLE-OF-NATION" STRATEGIC COMPETITION

by Christopher Ford, Marin Halper, and Andrea McFeely

MITRE's Center for Strategic Competition and the "Occasional Papers" Series

Today's competitive strategy challenges are multi-faceted and complex. They exist in arenas of hard, "sharp," and soft power—from military capacity, technological innovation, economic development, espionage, trade, and finance to cyber conflict, law enforcement, politics and culture, and diplomacy. Well-crafted competitive strategy requires a true system mindset.

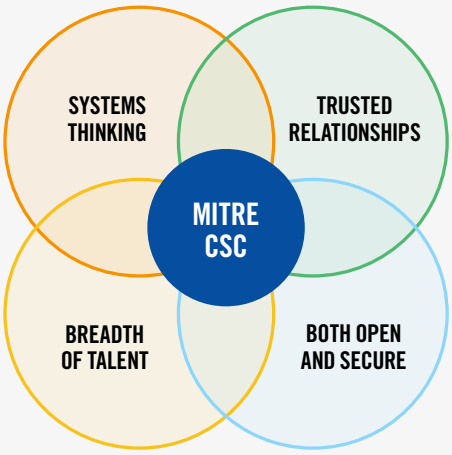
Helping our country meet these challenges drives the work of MITRE's Center for Strategic Competition (CSC). The CSC leverages MITRE's six decades of systems thinking and systems integration experience on behalf of the United States and its partners, drawing upon the full range of capabilities that exist across the MITRE enterprise.

Much of what we do occurs out of the public eye for specific federal sponsors. Because strategic competition is a genuinely "whole-of-nation" challenge, however, it is essential to involve a wide range of stakeholders in devising and implementing systems-informed solutions.

CSC established this "Occasional Papers" series to engage, educate, and inform the policy community and the broader public about strategic competition issues, problems, and opportunities. We invite you to send feedback to strategic.competitor@mitre.org.

Why do we need the Center for Strategic Competition (CSC)?

Understand complex interactions between elements of a system and devise interventions to improve outcomes



Decades of trusted relationships with uniquely broad range of federal sponsors, and in coordinating partnerships among government, industry, and academic stakeholders

Data analytics, economic modeling, AI, behavioral sciences, intelligence analysis, system-of-system modeling, and decision-support tools

Cutting-edge analytic capabilities integrating unclassified and classified information and analyses

MITRE's Mission

MITRE's mission-driven teams are dedicated to solving problems for a safer world. Through our public-private partnerships and federally funded R&D centers, we work across government and in partnership with industry to tackle challenges to the safety, stability, and well-being of our nation.

Contents

Executive Summary

U.S. leaders today find themselves facing an increasingly diverse array of "whole-of-nation" (WON) challenges—that is, problems that have broad, cross-cutting or even systemic effects, and that cannot be addressed by any given government department or agency, or even the federal government itself, acting entirely on its own. The Biden Administration, for instance, has called for America to respond on a WON basis to the problems of climate change, cybersecurity, and supply chain security, and that of strategic competition as a whole is surely also such a challenge. In response to such needs, the United States must figure out how to provide equivalently WON responses.

Unlike previous eras in which interest flared in WON policy implementation—such as when U.S. officials sought to establish cross-cutting policy "czars" in the 1990s or British ones spoke of "joined-up government"—we have the opportunity today to approach such challenges informed by data analytics more sophisticated than ever before. Finding policy answers to problems such as climate change, strategic competition, and security for the U.S. Defense Industrial Base (DIB) and national supply chains must be informed at all levels by state-of-the-art efforts to acquire, understand, and share data across traditional institutional "stovepipes."

A growing ecosystem of data aggregators, analysts, and artificial intelligence-enabled (AI) methodologies now exists to collect and help understand revealing patterns that may exist in the "digital exhaust" of the modern information economy—data which can provide the informational foundation upon which to build effective responses to such "wicked problems" at the national level. Massive datasets are now freely obtainable or commercially available on a fee-for-service or subscription basis, providing users with access to publicly available data sources around the world and across all of society and the modern economy. And increasingly powerful analytical and computational tools exist to assist with the "digestion" of such "nontraditional" data—an informal term denoting the fact that earlier generations neither really created nor certainly were able to acquire and aggregate it at scale. Aggregation of nontraditional data on this basis could thus potentially provide new insights into patterns and possible courses of action that would be invisible to traditional analysis. Such techniques are already employed in fields as diverse as public health, public safety, national security, economic development and poverty reduction, effective and accountable governance, education, and conservation and environmental protection.

It is not enough, however, for data to exist, or even for it to be accessed and analyzed. Once understood, relevant data must also be shared. In the context of WON organization, moreover, this sharing must occur across a range of traditional institutional and bureaucratic stovepipes. Nontraditional data analysis for such broad, cross-cutting purposes is thus only partly a challenge of technology. It is also a governance challenge: that of how to establish and maintain a regulatory, legal, and policy framework that permits effective data usage and sharing in support of WON strategy, but within constraints consistent with American values.

The category of "publicly available information" (PAI) is crucial to success in these regards, both because under current rules it is the category of information most sharable across the U.S. government, and because PAI makes up most of the nontraditional data presently available in the commercial marketplace. U.S. policy is to promote the effective use of PAI in support of policymaking and implementation. Nevertheless, certain aspects of PAI management—specifically, those related to information concerning "U.S. persons"—are subject to a set of overlapping federal regulations that provide concurrent, but inconsistent, standards to govern the handling and use of such data.

MERELY STICKING WITH THE *STATUS QUO* WOULD NOT PRECLUDE ALL EFFORTS TO UNDERTAKE DATA ANALYTICS TO SUPPORT POLICYMAKING IN THE FACE OF AMERICA'S "WICKED" PROBLEMS.

At least three such regimes presently exist: the framework created by the Privacy Act, rules established to govern the handling of such information for non-intelligence purposes at specific government agencies, and the U.S. intelligence oversight system. None of these rule sets entirely preclude aggregating nontraditional data and performing sophisticated analytical techniques on it in support of WON strategy, much less preventing more-subtle and less-intrusive methods whereby data is not acquired or stored in bulk by U.S. officials at all, but rather obtained via query-based access to federated networks of data repositories that already exist outside the government. The Privacy Act and the Executive Order (E.O.) 12333 intelligence oversight framework, in particular, are fairly permissive in allowing analysts to study nontraditional data at scale, provided this data meets the definition of "publicly available information." (Defense Department rules for accessing information about non-Defense-affiliated U.S. persons are more restrictive, but DoD permits modern data analytics in support of at least some tasks—such as securing the DIB and mitigating risks in U.S. supply chains.)

Nevertheless, the greatest national challenges the United States faces—among them climate change and strategic competition with China—involve cross-cutting substantive questions and policy issues in which a kaleidoscope of stakeholders must both understand their environment and act together effectively in support of some shared vision of American success. For this, the current governance model may not be enough. Merely sticking with the *status quo* would not preclude all efforts to undertake data analytics to support policymaking in the face of America's "wicked problems," but it would be inefficient, costly, and inadequate for the scope of today's challenges.

A better alternative, albeit a complex one, would be to build a nonprofit-run data center—a "trusted broker" and intermediary between public and private sector stakeholders—which would access publicly available nontraditional data—such as through query-based access to various repositories, with little reliance upon in-house aggregation or bulk storage. Sophisticated analytical techniques would then be used on this information, at scale, on behalf of participants in WON efforts, while adjusting this work to conform to the rule sets that currently govern data work on behalf of each user and mission requirement. This would require a management architecture that would allow the center to navigate existing governance frameworks on the fly, providing services to participating stakeholders on a "user-by-user" and "use-by-use" basis. (It would also likely require innovations in liability protection, resource-pooling, and institutional oversight.)

To help begin to provide such a "second-best" capability, the MITRE Corporation is already working to develop such an approach. An even better answer, however—or at least a complementary one—would be to reform federal data management rules on a system-wide basis to provide a uniform set of standards that apply across today's institutional boundaries, and under which institutional leaders, data aggregators, AI developers, auditing and oversight officials, and those charged with devising innovative use cases in support of WON missions could all work together to plan, build, and operate within a common set of well-understood parameters. Short of such full-spectrum reform, however, it might be possible to proceed incrementally, creating new, cross-cutting rules on an issue-by-issue basis. To the degree this issue-specific pilot program worked, these reforms could be expanded to additional areas. We would eventually reach full, system-wide harmonization, but the reform program would proceed step by step, with participants learning and adjusting approaches as they progress.

This paper suggests that such a pilot program begin with an effort to help secure America's critical supply chains against foreign adversary control or manipulation and offers suggestions for a new legislative framework that could accomplish the necessary harmonization in this arena. It also offers a tentative suggestion for a "Code of Ethical Conduct in the Use of Publicly Available Information" that could be promulgated as an articulation of best practices for the ethical employment of PAI.

Whatever the approach taken, however, it is essential we take full advantage of what sophisticated analysis of publicly available nontraditional data has to offer in helping meet our country's most pressing WON challenges. This will require access to massive volumes and breadths of data sources and analytical tools that are only now becoming available, and the U.S. system is not yet well-organized to permit this. Whatever the standard of privacy protection and institutional accountability that is adopted, it is essential we move toward a uniform standard for cross-jurisdictional data access, analysis, and dissemination in support of WON objectives. The White House, Congress, and other leaders should reform the federal data-management architecture in ways that facilitate WON competitive success, while protecting the values the American people cherish most.

Introduction

At a point in history when the United States faces formidable challenges that cut across substantive issue areas, across the jurisdiction of multiple federal agencies, across the divide between the public and private sectors, and across entire swathes of the America's economy and society—problems such as fighting and coping with the effects of climate change, or engaging in strategic competition with a "near-peer" adversary such as China—it has become almost commonplace for U.S. leaders to describe such problems as "whole-of-government" or "whole-of-nation" (a.k.a. WON) challenges. In response to these needs, it has thus become similarly commonplace to call for whole-of-government or whole-of-nation responses.

Just how the United States is to organize and equip itself for success in such work, however, is still less than clear. This paper, however, explores at least one element of this challenge: the need for sophisticated modern data analytics to support policymaking in developing and implementing genuinely WON strategies. In the pages that follow, we will outline the growth of nontraditional data aggregation and analytics in the private sector, the potential utility of such techniques in helping the United States address WON challenges, and the legal and regulatory terrain that U.S. officials will need to navigate—or perhaps to reform—to meet America's mid-21st-Century needs.

Modern Challenges and the Call for Whole-of-Nation Responses

WON: A Longstanding Aspiration

It is frequently observed that the United States faces multiple WON challenges today. The Biden Administration, for instance, has called for America to make WON efforts on several fronts. It has called for a "comprehensive, government-wide strategy to measure, disclose, manage[,] and mitigate the systemic risks climate change poses." It also says it is mounting a "whole-of-nation effort needed to address cybersecurity threats." Pursuant to President Biden's desire to "move the nation away from its reliance on adversaries—including China—for critical inputs," federal agency heads have also been directed to review U.S. supply chains and pursue mitigation steps all across government, and indeed in close consultation with outside stakeholders in private industry, academia, non-governmental organizations, labor unions, and state, local, and tribal governments.[4]

The "demand signal" for WON organization, policy development, and policy implementation is thus clear and strong. Our leaders recognize that they face WON problems, and they accordingly seek WON solutions.

To be sure, calls for such system-wide efforts have been made before. In the 1970s, for instance, multiple Western governments—confronted with challenges such as the Energy Crisis, stalled economic growth in the era of "stagflation," and newfound awareness of environmental pollution—explored the creation of "super ministries" intended to spur coordination across traditional governmental "stovepipes."[5] Such efforts went out of fashion during the anti-regulatory 1980s, but in the 1990s enthusiasm returned for such cross-cutting organizational forms. Most famously, the Labour government of former British Prime Minister Tony Blair advocated a concept it called "joined-up government." This idea was

> "based on the view that public policy goals cannot be met through the separate activities of existing organizations, nor can they be delivered by grouping several departments under a common agency. The idea is to align agency activities with particular goals, coordinating activities across organizational boundaries without removing the boundaries themselves."[6]

Such whole-of-government approaches—aimed at giving government some purchase on so-called "wicked"[7] problems such as terrorism, poverty, sustainable development, and pandemic risks, which seemed to defy ordinary approaches—were explored in several countries, including the United Kingdom, Canada, Australia, and even the United States.[8] The details of these various national efforts varied considerably, but they all tended to emphasize "the need for greater collaboration and coordination across departmental boundaries to eliminate duplication, optimize resources, [and] create synergies among agencies" to produce "coherent and integrated policies" supported by shared resources and seamless … communication, information sharing[,] and decision-making processes."[9] Today's calls for WON organization should be understood in light of such recurring aspirations for effective governance across economic sectors and traditional jurisdictional stovepipes.

But What Does It Mean?

As strong as the demand signal is today for WON strategies in response to challenges such as climate change and strategic competition, what is less clear is exactly how to organize and implement genuinely WON approaches in a democratic government. By contrast, it does seem fairly clear, at this point, what a WON strategy looks like in the *Chinese* context, where the Chinese Communist Party (CCP) possesses a wide range of administrative, regulatory, legal, and supra-legal coercive tools with which it can to some extent simply decree coordinated societal movement in any given direction.[10]

An approach to WON organization that is consistent with American values, however, would be notably different than the CCP's approach, for U.S. leaders neither have such heavy-handed coercive tools nor should ever be permitted to have them. As one of the authors has observed elsewhere, for instance, an analogue to Chinese practices is unavailable in the United States "because American leaders must not use government coercion to compel such cross-sectoral collaboration and hijack market mechanisms for state purposes."[11]

Nevertheless, though historical policies such as the British concept of "joined-up government" have sometimes seemed "rather vague about the actual mechanisms that would achieve that desired result,"[12] at least a few things seem clear. For one,

> "… [i]n finding answers that play to the strengths of our own political culture, the U.S. economy, and the free-market dynamism of our people, we must ensure that whatever forms we adopt to coordinate national efforts to catalyze innovation revolve around genuinely voluntary collaborations between governmental, private sector, and academic stakeholders …."[13]

WHOLE-OF-GOVERNMENT APPROACHES EXMPHASIZE COLLABORATION AND COORDINATION ACROSS DEPARTMENTAL BOUNDARIES, TO PRODUCE COHERENT AND INTEGRATED POLICIES SUPPORTED BY SEAMLESS COMMUNICATION, INFORMATION SHARING, AND DECISION-MAKING PROCESSES.

To make such voluntarism possible, WON organization also seems likely to require a good deal of "lower-level politics" in eliciting cooperative effort from diverse cross-sectoral stakeholders.[14] This, in turn, will surely entail

> "building a strong and unified sense of values, trust, values-based management, and collaboration; team building; involving participating organizations; and improving the training and self-development of public servants. There is a need to reestablish a 'common

> ethic' and a 'cohesive culture' in the public sector …. All agencies should be bound together by a single, distinctive ethos of public service."[15]

"Joining up" in pursuit of common goals across traditional institutional and sectoral divisions puts a special premium on information sharing and coordination, because "initiatives must align organizations with different cultures, incentives, management systems and aims, and they must align governments to citizens and their needs."[16]

Just what a full suite of best practices would look like in the development and implementation of WON strategies in a modern democracy, however, has not yet been fleshed out. As difficult as are the substantive questions about what substantive steps should collectively be taken to help solve problems such as climate change and strategic competition, genuinely WON organization *in response* to such wicked problems presents sociological, bureaucratic, cultural, political, and intellectual challenges of mobilization and coordination that are even more formidable still. This makes further research on the elements of effective WON organization essential.

Sketching out potential best practices in this regard is beyond the scope of this paper, however. Nevertheless, it is worth emphasizing here at least one of the elements that seems likely to be essential for any WON effort: the rapid and effective collection, aggregation, analysis, and dissemination of information. No WON approach can be effective if done in the dark; if policy answers are to be found for problems such as climate change and strategic competition, decision making must be informed at all levels by the best available data from across the traditional stovepipes of governmental and societal organization. Getting, understanding, and sharing data, in other words, will be critical to success.

The Supply Chain Example

To see the importance—and the challenges—of using data in support of a complex national strategy, for instance, one need look no farther than the difficulties of implementing President Biden's February 2021 Executive Order on securing America's supply chains.[17] In a complex modern economy, after all, a supply chain is "a gigantic puzzle consisting of various stakeholders and applications"[18] and stretching many layers deep. Typical approaches to understanding and mitigating supply chain risk try "to map out and assess the value chains of all major products," after which "[e]ach node of the supply chain—suppliers, plants, warehouses, and transport routes—is then assessed in detail."[19]

Yet illuminating such networks can be extraordinarily hard, because "most global supply chains are very complex[,] with hundreds of tier one suppliers and perhaps thousands of tier two and three suppliers."[20] Moreover, such data is "typically siloed and has to be gathered from multiple sources for aggregation,"[21] even where proprietary data restrictions do not prevent some of it from being gathered at all, "limiting visibility at the purchaser or integrating-manufacturer level." Accordingly, without extremely "robust processes to identify and successfully manage growing supply-chain risks as the world becomes more interconnected … supply-base transparency is hard (or impossible) to achieve."[22] Traditional bureaucratic methods of data collection and analysis are simply unequal to this task, and indeed below the first couple of layers of connection, even sophisticated supply-chain "primes" *themselves*—much less government officials—may have no idea from whence come many of the items and materials that go into key products. (Indeed, major contractors are generally *unable* to require such data from their own suppliers beyond their first tier, as subcontracting provisions seldom require such transparency, as a result of which prime contractors are excluded from awareness due to contract privity.)

As the supply chain example suggests, the sort of wicked problems that are so often the focus of WON policy approaches are, in important ways, *data* problems. If soluble at all, they most certainly cannot be solved without much more—and better—data than our system

of governance is traditionally capable of acquiring, handling, and understanding.

To be sure, this is not to suggest that merely collecting and understanding more data about such challenges is in itself sufficient to permit their solution. To the contrary, wicked problems are frequently—from a policymaker's perspective—greater than simply the sum of their parts. Even the best data access and analytics, for instance, cannot in themselves solve the challenges of organizational behavior, cross-sectoral social mobilization, and management science that confront us in implementing WON strategies. Nor is simply getting "more and better data" a sufficient answer to the decision-making and leadership challenges presented when large numbers of diverse, interconnected, and reciprocally influencing individual and organizational entities come together into "complex adaptive systems."[23] (Such systems behave in nonlinear ways influenced by both positive and negative feedback loops and sometimes give rise to "emergent" higher-order behavioral patterns. These dynamics problematize traditional notions of linear government policymaking in which purposive policy inputs are presumed to lead predictably to specified situational outcomes.)

Yet even the most sophisticated and wisely implemented policies can do little without data. Furthermore, due to their cross-cutting nature, today's WON challenges are likely to be especially—and inescapably—data-*intensive*, requiring not simply access to an unprecedented *volume* of data but also a sprawling *diversity* of data sources.

But how are modern policymakers to handle such extraordinary data demands?

As the supply chain example suggests, the sort of wicked problems that are so often the focus of WON policy approaches are, in important ways, *data* problems. If soluble at all, they most certainly cannot be solved without much more—and better—data than our system of governance is traditionally capable of acquiring, handling, and understanding.

An Era of Big Data

Data and Its Challenges

Fortunately—and in sharp contrast to periods in the 1970s and 1990s in which governmental enthusiasm for WON organization has bloomed—there exists today no shortage of data. Indeed, in some respects there would seem to be so *much* data available that the greatest challenge now lies in figuring out how to make sense of the torrent of it constantly produced as the "digital exhaust," as it were, of the modern economy.[25]

A growing number of data aggregators, analysts, and AI-enabled services, however, believe they can. Hoping to gain competitive insights into market behavior, for instance, hedge funds have invested heavily in acquiring and seeking to understand the "nontraditional data"[26] constantly being generated as a byproduct of all manner of commercial transactions. As the range of potential data use cases and the budgets of data purchasers have soared in recent years, so too have the number of aggregators providing such data as a commercial service—already reaching into the multiple hundreds even by 2018.[27] One survey that year, in fact, estimated that spending on such services by institutional asset managers, hedge funds, and proprietary trading firms totaled some $300 million a year, having doubled from the year before.[28]

With a range of diverse data sources coming to be routinely drawn upon to inform all manner of investment and business decisions, data aggregation is today one of the fastest-growing sectors in the world,[29] and the annual global market for such analytics has been predicted to reach a remarkable $132 billion by 2026.[30] According to one entrepreneur, at least, "Big Data is the new oil. The companies, governments and organizations that are able to mine this resource will have an enormous advantage over those that don't."[31]

There is undoubtedly at least some hype in such predictions of the devastating competitive advantages

supposedly available through the use of data analytics, for the commercial providers of data aggregation services have no small incentive to promote the idea that the analysis of nontraditional data can accomplish miracles for their customers. Some observers also worry that users will place undue confidence in the results of such analysis—especially where the prediction of complex phenomena is attempted, rather than simply the illumination of subtle patterns and connections—or that unscrupulous users will attempt to "game" data analytics and skew its results,[32] or that some might use data-aggregation techniques to infringe upon citizens' privacy.[33]

On the whole, however, expert respondents surveyed on the subject have seemed more optimistic than pessimistic,[34] and undoubtedly, massive data aggregation and analysis has become enormously widespread and is now used to support a growing number of decision-support use cases. Such tools are certainly not going away, and indeed they are expanding both in their analytical power and scope, with ever-more-diverse sources becoming available and improvements being made in the artificial intelligence (AI) algorithms and automation used to sort and to explore correlations within such data.[35]

From the perspective of WON strategy, these developments highlight the importance of taking advantage of what this field has to offer in helping U.S. leaders address massive, cross-cutting national challenges. They also highlight the importance of doing so within a framework of privacy protections and oversight accountability that is consistent with American values. The Big Data[36] era is here to stay, however, and if we aspire to genuinely WON answers to our biggest problems, *failing* to take advantage of nontraditional data analytics is not an option.

Booming Data Sources and Use Cases

The range of data available to support such analytics is indeed extraordinary. Many data sources are available openly on the internet, where even a quick search can find websites dedicated to providing users with long lists of resources that include government and political data, social media databases, weather data, sports data, news data, university and research center data, consumption data, health data, economic data, and the products of a range of openly available data aggregator sites.[37] Such datasets are frequently free for the taking, essentially being offered as an information-age public service for the common good. Also publicly available, however—albeit generally on a fee-for-service or subscription basis, to compensate aggregators for the trouble they take to compile it—is data collected through customized or proprietary search methods devised by specialist companies that build web-crawling "bots" to collect, deduplicate, and normalize it for their customers.[38]

Altogether, the range of freely or commercially available data sources stretches around the world and across all of society and the modern economy. Data analysis done today in the private sector, for instance—such as by the aforementioned hedge funds and other financial and commercial players—already draws upon sources such as smartphone application installations and usage, credit and debit card transaction and point-of-sale data, e-mail and consumer receipts, public data, satellite data, social media information, survey data, weather data, data "scraped" from public websites, web traffic data, search engine trends data, crowd-sourced data, business performance metrics, shipping manifests, ocean vessel tracking information, logistics data, commercial "footfall" and geolocation data, business transactions, and more.[39]

Such analytics could potentially provide new insights into patterns and corrections that would be invisible to traditional analysis. As noted, the cutting edge of this field seems to lie in doing such nontraditional data analysis for profit, in search of competitive advantage in finance or commerce.[40] But such tools may also offer insights to help solve broader societal problems—such as in public health, public safety, national security, economic development and poverty reduction, effective and accountable governance, education, and conservation and environmental protection.[41]

Healthcare, for example, already benefits in important ways from data analysis at scale, for this can permit better monitoring of healthcare needs and service quality, helping identify opportunities to improve care for underserved populations, providing higher-quality analysis of treatment efficacy and outcomes, and supporting the development and deployment of improved medications and treatments. It was recognized years ago that healthcare can be improved

> "through information sharing and technology, leading to a focused, knowledge-based healthcare system. By providing data aggregation, data warehousing, performance management and analytics, healthcare executives can arrive at deeper insight. They can devise more effective strategies to increase quality and control costs."[42]

In this field, in other words, it is often the case that "more data equals better healthcare."[43]

In search of further such insights, the National Science Foundation and the National Institutes of Health have been cooperating for years "to develop new methods to derive knowledge from data[,] construct new infrastructure to manage, curate and serve data to communities[,] and forge new approaches for associated education and training."[44] MITRE—for which, by way of full disclosure, the authors of this paper all work—is also today engaged in a partnership to develop and pilot the use of foundational cancer data elements collected from electronic health records through an open-source, nonproprietary model for data interconnectivity, to allow aggregation of standardized information sets with data from many other sources and analyzed for oncological best practices. Participants in this "minimal Common Oncology Data Elements" (mCODE™) project have different roles—some training AI algorithms on large data sets and others focusing upon particular data sets to generate mCODE elements, while MITRE focuses on ensuring quality, trust, and provenance of the data most needed at a cancer patient's point of care—but the whole effort centers on the idea of improving health outcomes by "accelerating data sharing and aggregation and by the creation of a learning health system in which routine patient care data seamlessly inform scientific discovery, and, reciprocally, research informs practice."[45]

In the security realm, there has been considerable interest for years in how broad data access and pattern analysis, including the use of AI algorithms, might improve counterterrorist intelligence analysis.[46] Analytical tools and information fusion techniques involving social network analysis of urban gangs, citywide alert systems, crime-spot prediction, and custody decision-making aids are also "already in use by law enforcement agencies" in a number of countries.[47] (Nor are our governmental "good guys" alone in such endeavors, for as we have learned in recent years, America's adversaries—unconstrained by our oversight rules and ethical standards—are also learning to analyze diverse data sources, including social media activity, such as in targeting propaganda and disinformation efforts at specific U.S. audiences.[48] In a thriving data marketplace, the "bad guys" can purchase insight as easily as anyone else, and they appear to be doing so.)

This expanding range of Big Data use cases adds weight to the intuition that if we are to gain any purchase on cross-cutting WON problems such as climate change and strategic competition, the U.S. government will need to be much better at accessing and analyzing nontraditional data than it has been to date. To return to our earlier example of President Biden's February 2021 Executive Order on securing America's supply chains,[49] for instance, it is likely that the remarkable breadth and depth of the data now publicly available from the digital exhaust of the modern economy can be used to glean important insights into the supply chain—and into risks associated with its dependence upon or subversion by great power adversaries—that would be quite invisible to traditional methods.

Quite a bit of such "exhaust" is generated on a routine basis by the ordinary activity of corporate and business relationships and commercial transactions, and this can be a potent source of supply chain insight. It is already the case, in fact, that commercially available business intelligence tools can help identify otherwise-hidden relationships between corporate entities.

One study published in 2017, for example, analyzed the contents of a publicly available company ownership database that then covered about 200 million public and private firms worldwide, and used this data to identify the operating revenue, country, city, sector, global ultimate owner, and all ownership relationships (and direct and total ownership percentages) for each available company. This work revealed "71,201,304 distinct ownership relationships between 98,255,206 companies," permitting the authors of the study to sort companies into "global ownership chains" in ways that let the analysts illuminate many of the otherwise highly obscure ways in which multinational corporations use complex corporate structures of parents and subsidiaries to organize their global operations and management.[50]

Such work—along with other published studies that use large-scale data analysis to identify important hidden trends and dynamics[51]—hints at the ways in which nontraditional data analytics of the world's digital exhaust could powerfully illuminate supply chain relationships and help the U.S. government mitigate supply chain risks arising from obfuscated corporate ownership relationships, nefarious actors posing as domestic companies, unobvious foreign-owned or -held companies, or those with undisclosed foreign partnerships or interests. More broadly still, modern data analytics likely have great promise in facilitating supply chain risk management, in conducting contracting and other private sector due diligence—including for purposes of technology-transfer "de-risking"[52]—in law enforcement and public integrity investigations, in implementing national-security export controls, in screening foreign investments for security risks, and in intelligence analysis.

The Cross-Jurisdictional Challenge

A Policy and Governance Challenge

But for the blizzard of nontraditional data available today to be most valuable in WON endeavors, it is not enough for such data merely to exist, or even for it to be painstakingly accessed and carefully analyzed using state-of-the-art techniques. Once understood, relevant data, or discovered outputs from the data, must also be *shared*. In the context of WON organization, moreover, this sharing must occur across a range of traditional institutional and bureaucratic stovepipes. It will surely not be easy to do this, especially at first, for our system of governance is traditionally ill-suited to such broad sharing. The need for such sharing, however, is inescapable for any genuinely WON effort to meet a WON problem. For success, accessing data, understanding it, and implementing effective information sharing based on the results all need to be possible—and indeed actually to occur—across organizational silos.

Even in data-analytical contexts and multi-stakeholder environments less institutionally challenging than those involved in implementing WON strategies—such as in the analysis of terrorist radicalization—it has been recognized for some years that *information fusion* is a critical ingredient of success. One counterterrorism-focused RAND Corporation study in 2013, for instance, stressed repeatedly that it is possible to combine "machine-learning and big-data analysis … [to] 'discover' unknown patterns or activities hidden in large amounts of data" with the fusion of information coming from a wide variety of sources[53] In the context of data aggregation and analytics, however, this can be very difficult, for disparate nontraditional data sources may provide a mix of quantitative and qualitative information without an obvious mechanism for combining them.[54] Nonetheless, RAND experts warn, while information fusion is "not a panacea," it "seems to be the only hope."[55]

In Big Data analytics, the more one is able to put in, as it were, the more one potentially stands to get out of the process: "More data builds better models."[56] Yet—in contrast to the comparatively "Wild West" environment of private-sector data aggregation by hedge funds or technology hyperscalers such as Google or Meta (a.k.a. Facebook)—our society does impose some limits and restrictions upon acquiring access to and sharing data in support of government activity. This challenge has been seen as particularly acute in counterterrorism, where as Kathleen McKendrick has noted,

> "[t]he quality of models achievable is limited by restrictions on what types of data can be accessed and how those data can be used. For government agencies, limits are imposed based on national regulations, international human rights laws or physical access and technical capabilities. … Paradoxically, restricting access could limit the ability to develop good models that could otherwise improve compliance with conditions of proportionality and non-discrimination."[57]

The basic challenge of how to do mission-critical data analytics at scale in a privacy-protective society, however, is a more general one. It is for this reason "the increasing availability of … [publicly available] information, including in social media," has been described as likely to "be among the most significant challenges and opportunities for U.S. intelligence and counter-intelligence in the coming years."[58] How to devise effective and appropriate rules and procedures for such work is a key challenge for WON organization.

A Critical Ingredient: Publicly Available Information

The category of publicly available information (PAI) is crucial to success in these regards, both because under current rules it is the category of information most sharable across the U.S. government, and because PAI makes up the vast majority—albeit perhaps not all—of the nontraditional data presently available in the commercial marketplace for Big Data analytics. In effect, this makes the effective access, analysis, and cross-institutional sharing of PAI a *sine qua non* test case for WON organization in the information age, for if the government cannot do such data analytics and information sharing effectively even with PAI, it seems hard to see how it could be done with any other sort.

Various sets of rules and regulations cover what different U.S. government agencies can do by way of acquiring[59] and using information, but all federal definitions of PAI follow the same pattern. For overall DoD purposes, for instance, PAI is defined as

> "Information that has been published or broadcast for public consumption, is available on request to the public, is accessible online or otherwise to the public, is available to the public by subscription or purchase, could be seen or heard by a casual observer, is made available at a meeting open to the public, or is obtained by visiting a place or attending an event that is open to the public."[60]

Despite slight differences in phrasing, the formulation used by the U.S. Intelligence Community (IC) parallels this DoD definition. The U.S. Attorney General's September 2008 *Guidelines for Domestic FBI Operations*, for instance, define PAI as

> "Information that has been published or broadcast for public consumption, is accessible on-line or otherwise to the public, is accessible to the public by subscription or purchase, could be seen or heard by any casual observer, is made available at a meeting open to the public, or is obtained by visiting any place or attending any event that is open to the public."[61]

This standard is important from the perspective of hedge-fund style data integration for purposes of implementing WON strategy, for as these definitions suggest, information will be PAI if it is available by purchase or subscription to any member of public

without further vetting by provider.[62] And if, in turn, information *is* PAI, even members of the IC—the portions of the U.S. government arguably subject to the sharpest restrictions on handling data that could potentially include information about Americans—can generally possess it. As discussed in more detail below, for example, under E.O. 12333, IC components are permitted to collect, retain, and disseminate "information that is publicly available"[63] provided there exists "a valid mission requirement" for using that information.[64] (Under this rubric, in fact, the IC has developed the sub-specialty of open-source intelligence—or OSINT—which involves "gathering, analyzing, and interpreting publicly available data" such as media information, internet content, public government data, professional and academic publications, commercial data, and the like.[65])

Significantly—and of great potential import for implementing WON strategies in the face of wicked problems such as climate change and strategic competition—it is a point of emphasis for U.S. policy in both DoD *and* the IC to promote widespread and effective use of PAI. This policy focus is especially clear in the Defense Department, which in 2019 issued its DoD Directive 3115.18 in large part precisely to

> "elevate[] and operationalize[] the use of publicly available information for the Department of Defense to users outside of the traditional open source, intelligence and investigative realms. That means PAI is no longer strictly the concern of intelligence analysts, cyber analysts and investigators."[66]

That directive makes clear that "DoD may access, obtain, and use PAI to plan, inform, enable, execute, and support the full spectrum of DoD missions"[67] and that DoD components "will have the ability to access PAI relevant to their missions" through appropriate data management and dissemination systems.[68] (DoDD 3115.18 even instructs departmental personnel to "*share* PAI and PAI tools with federal, State, local, tribal, and foreign partners" in accordance with applicable rules.[69])

The policy emphasis on effective utilization of PAI seems less strong within the IC, which has sometimes been described as having a congenital—and "deadly"—bias toward reliance on classified sources, sometimes almost to the point of disdaining open-source analysis, and which is said to have shown "reluctance to use [its] authorities" to collect, analyze, and disseminate OSINT.[70] Nevertheless, even the IC claims today to place a "heightened emphasis on reaching out externally for expertise to inform analysis,"[71] and the U.S. Defense Intelligence Agency (DIA) is working to revamp and make more effective how the IC gathers and analyzes data from social media and commercially-available datasets.[72]

Clearly, better data access and analytics are needed, both within the IC and—crucially, from a WON perspective—more broadly across the U.S. government in support of WON strategy. According to some reports, for instance, as much as 80 percent of DIA intelligence reporting now derives from unclassified sources, and "the military [is] ingesting so much data—much of it from unclassified open sources—that processing and analytical power [a]re now in short supply."[73] Meanwhile, PAI-based data-acquisition and decision-support analytics outside the IC are apparently in an even more parlous and merely embryonic state, making reform in this area an important national priority.

Whether for purposes of tracking WON climate change mitigation and response policies, organizing the U.S. system for strategic competition against China, securing American supply chains and protecting the U.S. DIB, or implementing nationwide "Horizon Strategies" to reinvigorate the American Innovation Economy,[74] we need to be much better at acquiring, understanding, and sharing insights gained through the analysis of nontraditional data at scale. Innovative and effective uses of PAI—including analysis of the digital exhaust discussed above—will be critical to our success in these areas.

Overlapping Frameworks

One factor that has slowed the U.S. government's ability to keep pace with the private sector in using PAI in trying to address national-level problems is that aggregations of PAI from the everyday operations of the global economy likely contain at least some information about ordinary Americans: "U.S. persons," in technical jargon. (This is not a challenge that impedes America's adversaries, however, who are presumably as free to purchase data aggregation services in the marketplace as anyone else, and who, as we have seen, are now beginning to do so.) This creates some policy, legal, and regulatory challenges for large-scale data analytics.

Precisely because what is at issue here is merely *publicly available* data—which by definition exists in *public*, and that literally anyone could in theory freely acquire—the sensitivity of PAI is minimal. Nevertheless, the modern U.S. system is extremely attentive to Americans' privacy rights, at least where the government itself is concerned. (Hyperscalers and hedge funds have long gathered, for their own profit, far more about U.S. persons, and in more detail, than the U.S. government will likely ever collect, and such aggregation is often actually essential to their business models.) U.S. court cases have also suggested that at least in extreme cases—such as when law enforcement agents obtain cell tower data revealing all of an individual's movements—acquiring revealing data about Americans will run afoul of the Fourth Amendment's constitutional prohibition of unreasonable searches and seizures (and thus require a search warrant), even though that information had already been provided to a third party by the individual in question.[75] Congress has also used legislation to regulate how information about Americans can be acquired and handled by government officials.

But here lies the difficulty, for there now exist *multiple* legal and regulatory regimes for controlling the acquisition and use of information that might contain data about U.S. persons. These frameworks were created at different times and for different purposes, but their subject matter is similar and they overlap to a considerable degree, with each imposing somewhat different requirements and with no one yet having harmonized and rationalized their concurrent strictures. U.S. government work to take advantage of the policymaking opportunities offered by nontraditional data analytics using PAI, in other words, needs to be able somehow to navigate this patchwork of regimes seamlessly in the course of day-to-day operations, and this is no mean feat. The following pages will outline three such regimes: the framework created by the Privacy Act,[76] DoD rules established to govern the handling of such information for non-intelligence purposes, and the U.S. intelligence oversight system that applies under E.O. 12333[77] and agency-specific implementing regulations related to U.S. person information (USPI).

The Privacy Act Framework

The Privacy Act of 1974 was—as the U.S. Justice Department archly notes—"passed in great haste" by Congress during the final week of a legislative session, and without the usual expedient of a conference committee having reconciled inconsistencies passed in separate versions of the bill by the Senate and the House of Representatives.[78] Politically and substantively, it represented a visceral legislative reaction to "a crisis of public trust" in government, including to revelations about federal surveillance of supposedly "subversive" elements in the American population.

> "Enacted in the wake of the Watergate and the Counterintelligence Program (COINTELPRO) scandals involving illegal surveillance on opposition political parties and individuals deemed to be 'subversive,' the Privacy Act sought to restore trust in government and to address what at the time was seen as an existential threat to American democracy … [by placing] 'limits upon what the Government can know about each of

> its citizens.' … In drafting the Privacy Act, Congress [also] relied on a recently published and widely read report from an advisory committee … [on] the risks to privacy presented by the increasingly widespread use of electronic information technologies by organizations, replacing traditional paper-based systems of creation, storage, and retrieval of information."

The Privacy Act regulates how most government agencies can handle information about American "individuals"—that is, U.S. citizens and aliens lawfully admitted for permanent residence, by specifying how agencies must handle such records. The statute defines a record as

> "any item, collection, or grouping of information about an individual that is maintained by an agency, including, but not limited to, his education, financial transactions, medical history, and criminal or employment history and that contains his name, or the identifying number, symbol, or other identifying particular assigned to the individual, such as a finger or voice print or a photograph."

The Privacy Act's core rule is that no government agency may disclose any record contained within a "system of records" to anyone, without the consent of the individual in question, except under specified conditions—such as where the recipient is an official in the agency maintaining that record who has need of it in the performance of official duties, or for "routine use" by a government agency.[82] A "system of records" means "a group of records under the control of any agency from which information is retrieved by the name of the individual or by some identifying number, symbol, or other identifying particular assigned to the individual."[83] "Routine use" means the use of that record "for a purpose which is compatible with the purpose for which it was collected."[84] In sum, though the term "disclose" is not actually defined in the Privacy Act,[85] a government agency is generally barred from sharing specific information about an individual citizen or lawfully resident alien out of any system the agency queries for this information on the basis of a specific "identifying particular" associated with that individual, for any purpose other than that for which the information about that person was collected in the first place.[86]

To be sure, some idiosyncrasies exist in this system, which may leave at least some scope for data analytics at scale notwithstanding the Act's general restrictiveness. For one thing—although, as we explore further below, the rules in this respect differ under the U.S. intelligence oversight framework—the Privacy Act applies exclusively to individuals and offers no specific protection for U.S. *companies* or other collective entities made up of citizens or lawfully resident aliens. Moreover, an electronic record-keeping system only counts as a system of records under the Privacy Act if information actually is retrieved therefrom on the basis of some identifying particular associated with individual persons. (It is not enough, in other words, that such a database is capable of responding to individually specific queries: it must in fact be used in this way to be covered by the Act.)[87]

In addition, under implementing regulations—at least at DoD, at any rate—the entire Department is considered a single entity for Privacy Act purposes. As a result, sharing information *within* the Department (i.e., between constituent components) is permitted where "[t]he requester has a need for the record in the performance of his or her assigned duties."[88] This also applies to contractors working for a government agency, as they are considered part of that agency[89] and may both receive and share information as if they *were* that agency. DoD regulations, for example, specify that disclosure of information by a DoD component

> "to a contractor for use in the performance of a DoD contract is considered a disclosure within the Department of Defense. … The contractor is considered the agent of the contracting DoD Component and to be maintaining and receiving the records for that Component."[90]

It would thus appear that (1) there is no Privacy Act problem with acquiring and sharing information about any company or collective entity, (2) aggregating and sharing information resulting from queries that are *not* made on the basis of specific personal identifiers of an American citizen or lawfully resident alien are permitted under the Privacy Act even if such queries *return* information specifically about such a person, and (3) various components *within* DoD and contractors employed by the Department can share information among themselves for legitimate official purposes, without raising Privacy Act concerns. (It would also be possible to do more than that, of course—such as creating a database of fused nontraditional PAI data to analyze without restriction in support of tasks such as securing the DIB, imposing sanctions and export controls, identifying supply chain risk, tracking climate policy implementation, and performing "net assessments" of the state of technology competition with China. In such case, however, one would have to comply with Privacy Act rules such as formally publishing a statement of records notice—or SORN—in the *Federal Register* declaring the existence of such a database and the routine uses to which it would be put.[91]) These wrinkles may thus make some modern PAI-based data analytics possible, at least if one is careful.

Non-Intelligence Information Collection

Information about Non-DoD Individuals and Organizations

Additional frameworks exist in the U.S. federal system that bear upon whether, when, and how agency personnel can acquire and share even PAI that might contain data about Americans. Such concurrent frameworks—which exist in addition to Privacy Act rules and the intelligence oversight system—increase the complexity of the regulatory terrain that must be navigated if the United States is ever to achieve a genuinely WON response to a challenge such as climate change or strategic competition with China.

One of these frameworks is provided by DoD Directive 5200.27, which regulates the "Acquisition of Information Concerning Persons and Organizations not Affiliated with the Department of Defense" and applies to DoD components *outside* the IC. By its own terms, this Directive covers "the acquisition of information concerning the activities of" persons not affiliated with DoD within the United States, as well as "non-DoD-affiliated U.S. citizens anywhere in the world."[92] As a general rule, it declares, Defense Department policy "prohibits collecting, reporting, processing, or storing information on individuals or organizations not affiliated with the Department of Defense."[93]

However, exceptions exist for "limited circumstances where such information is essential to the accomplishment" of specific DoD missions.[94] These exceptions relate to "activities threatening defense military and civilian personnel and defense activities and installations, including vessels, aircraft, communications equipment, and supplies."[95] This list effectively describes *functions* for the performance of which collecting information about non-DoD-affiliated U.S. persons is permitted. Notably, public availability is irrelevant here: all that matters is whether it concerns a non-DoD affiliated U.S. person and *why* some DoD component seeks to collect information.

Most of the activities described—data collection for which is thereby permittable—relate to direct physical or organizational threats to the DoD or U.S. armed forces personnel. They thus allow information to be collected related to such things as subversion, sabotage, theft, and other lawbreaking, as well as for purposes of personnel security investigations and operations related to civil disturbances.[96] Most of these categories, therefore, are not obviously relevant to the challenges of organizing the U.S. government for—and equipping it with the information and insight needed to implement—WON strategies.

Of the list provided in the Directive, however, two may have some significance here. The Directive specifies that "[a]cts jeopardizing the security of DoD elements or operations

or compromising classified defense information by unauthorized disclosure or by espionage" and "[a]ctivities endangering facilities that have classified defense contracts or that have been officially designated as key defense facilities" are among those about which information may be collected by DoD personnel about non-DoD-affiliated U.S. persons.[97] These two stipulations may be difficult to relate to some WON problems, of course, but it would not be unreasonable to think that these provisions would indeed permit DoD to collect information—and engage in data aggregation and analysis at scale—on potential threats to the DIB and to military supply chains. Such threats, for instance, could include penetration by foreign entities, the manipulation of strategic dependencies by foreign adversaries, or the insertion of faulty or doctored components or materials. Beyond this, however, the Directive would seem a significant impediment to the kind of data-driven policy support we have been discussing.

Federal Acquisition Regulations

There may be additional opportunities to use PAI in helping secure American supply chains in connection with DoD efforts to mitigate adversary Foreign Ownership, Control, or Influence (FOCI) in defense-critical supply chains and to safeguard markets. The Department recognizes that this work will require a solid analytical foundations, rooted in new efforts to gain visibility into and to understand those supply chains:

> "The Department is committed to protecting its supply chains and the defense industrial base from adversarial FOCI by scaling efforts to identify and mitigate FOCI concerns. … This effort requires a front-end assessment of a program's acquisition strategy to ensure a resilient supply chain. Early identification of any FOCI concerns enables mitigation before contract or grant awards. The Department will scale its efforts to identify and mitigate FOCI in supply chain decision-making to ensure investments are not degraded through counterfeit, compromise, or theft."[98]

And indeed, there may already exist considerable authority to do this. The Federal Acquisition Regulations (FAR), for instance, allow—or, more specifically, actually direct—government agencies to use market research to search for sources capable of satisfying agency requirements and to ascertain the practices of firms engaged in producing, distributing, and supporting relevant commercial products or commercial services.[99] The techniques government agencies may use in such research include "[q]uerying … Government and commercial databases that provide information relevant to agency acquisitions."[100] With DoD requirements focusing with ever-greater specificity upon the need for a "robust and secure … industry and supply chain" to provide mission-critical items and materiel and upon the need to "[i]dentify FOCI and other supply chain risks,"[101] these FAR market research authorities may permit quite a bit of PAI-informed data-analytical work to be done to understand the national security supply chains and the nature and practices of the entities therein.

Intelligence Oversight Framework

Basic Rules

In general, as noted earlier, the intelligence oversight framework established by E.O. 12333 permits the acquisition, analysis, and sharing of PAI. Under that Order, IC elements shall be permitted "collection, retention, and dissemination of … [i]nformation that is publicly available."[102] IC components, moreover, are permitted to "enter into contracts or arrangements for the provision of goods or services with private companies or institutions in the United States," and when doing so, such components "need not reveal the sponsorship of such contracts or arrangements for authorized intelligence purposes."[103] All this enables intelligence agencies to acquire PAI from commercial aggregators on the open market and to apply analytical techniques to it in the various ways we have been discussing.

Nevertheless, components of the U.S. IC also fall under a set of rules related to handling information about U.S. persons—rules that are entirely separate from the Privacy Act but impose restrictions that apply concurrently. E.O. 12333 defines a "United States person" as

> "a United States citizen, an alien known by the intelligence element concerned to be a permanent resident alien, an unincorporated association substantially composed of United States citizens or permanent resident aliens, or a corporation incorporated in the United States, except for a corporation directed and controlled by a foreign government or governments."[104]

Notably, this is broader than the class of entities protected by the Privacy Act, in that the IC's "U.S. Person" rule applies also to groups: associations substantially composed of U.S. citizens or permanent resident aliens and U.S. corporations.

The specific procedures for how to handle USPI are provided on an agency-by-agency basis.[105] According to E.O. 12333,

> "[e]lements of the Intelligence Community are authorized to collect, retain, or disseminate information concerning United States persons only in accordance with procedures established by the head of the Intelligence Community element concerned or by the head of a department containing such element and approved by the Attorney General …."[106]

The following pages will therefore discuss two illustrative examples of these agency-specific rules—for the Central Intelligence Agency (CIA), and for the intelligence components of DoD—from the perspective of aggregating, analyzing, and sharing nontraditional data at scale.

Agency-Specific Implementing Rules

Central Intelligence Agency

With regard to how it protects the privacy of U.S. persons in handling PAI, the CIA operates under a set of guidelines approved by the U.S. Attorney General, and which were declassified and released in 2017.[107] Not surprisingly, these rules are based upon the definition of a U.S. person contained in E.O. 12333, though they do add some potentially helpful clarifications—such as in making clear that "[a]n alien who procures a visa or other documentation by fraud or willful misrepresentation of a material fact is not a lawful permanent resident for purposes of these Procedures," and by setting forth a rule of presumption whereby someone known to be inside the United States is presumed to be a U.S. person (and someone outside the country is presumed not to be) until specific information is obtained to the contrary.[108]

The CIA procedures are particularly concerned with what is termed "United States Person Identifying Information" (USPII), which the CIA must generally remove from whatever data it acquires before it can be retained and disseminated.[109] "To the extent practicable," the CIA is enjoined to remove all USPII before dissemination of any information outside the IC—e.g., to the policymakers who are key consumers of the Agency's intelligence reporting—"unless it is necessary or reasonably believed that the information may become necessary to understand, assess, or act on the information being disseminated."[110]

The category of USPII, in turn, is defined under the CIA procedures as information that is

> "reasonably likely to identify one or more specific U.S. persons. USPII may be either a single item of information or information that, when combined with other information, is reasonably likely to identify one or more specific U.S. persons."[111]

In this sense, USPII is a subset of "information concerning U.S. persons." (It is, however, a somewhat indeterminate one. The CIA procedures note that identifying USPII "in a particular context may require a case-by-case assessment by a trained intelligence professional. It is not limited to any single category or information or technology."[112]) The CIA's "default rule," as it were, is that USPII may not be retained or disseminated.

Nevertheless, there is for present purposes a very important exception. Under the CIA procedures, the Agency may retain information that is publicly available, *even if it contains USPII*.[113] Nor is the CIA merely permitted to retain publicly available USPII it may have collected "incidentally" to other lawful intelligence collection.[114] It may also deliberately *collect* information concerning U.S. persons, provided this information is publicly available. This is quite explicit, for the procedures specify that the Agency may "collect, retain, and disseminate" PAI concerning U.S. persons when this is done "in the course of CIA's duly authorized intelligence activities and in fulfillment of the CIA's national security responsibilities."[115] (Collecting PAI concerning U.S. persons, moreover, is considered no more than "basic collection,"[116] which may be done by a CIA employee—for an authorized official purpose, of course, and only to the extent reasonably necessary to support that purpose[117]—without any special approval from supervisors.[118])

So what counts as PAI? In line with E.O. 12333, the CIA procedures define it as

> "information that has been published or broadcast for public consumption, is available on request to the public, is accessible online or otherwise to the public, is available to the public by subscription or purchase, could be seen or heard by any casual observer …, is made available at a meeting open to the public, or is obtained by visiting any place or attending any event that is open to the public. Information is publicly available only if it is made available to the CIA under conditions or on terms generally available to the public."[119]

All of this, then, is consistent with the acquisition, analysis, and sharing of nontraditional data, even if such data contains information about U.S. persons. In light of the burgeoning ecosystem of commercial aggregators and suppliers of nontraditional data that has developed in recent years, it is also significant that the CIA procedures expressly note that "certain commercially acquired data may be considered publicly available if a non-U.S. government person or corporation could acquire the same data in that same way from that same commercial source."[120]

In general, therefore, as summarized by the former head of the Justice Department's National Security Division, under these CIA procedures

> "[p]ublicly available information, including information concerning U.S. persons, may be retained indefinitely (subject to NARA records control schedules) even if it contains USPI[I]. The retained information generally may be queried 'if the query is reasonably designed to retrieve information related to a CIA authority and responsibility.' Information concerning a U.S. person may be disseminated freely within the CIA, and to another IC element if relevant to that element's responsibilities (or for the purpose of determining whether the information is relevant). Publicly available information concerning U.S. persons, including USPI[I], also may be disseminated freely."[121]

To be sure, special procedures are provided for where it has not yet been possible to evaluate collected information to determine its status (e.g., whether it contains USPII). Such "unevaluated information" can only be queried for a duly authorized CIA activity under limited circumstances, such as when that query is "not designed to retrieve information concerning a U.S. person," or at least when such query is "to the extent practicable … accompanied by a statement explaining the purpose of the inquiry."[122] (Unevaluated information is also subject to special procedures when it comes to disseminating such information outside the CIA.[123]) Special rules also cover "bulk collection"—which is defined as "the collection of data that, due to technical or operational considerations, is acquired without the use of discriminants (e.g., specific identifiers, selection terms, etc.)"[124]—with the result that specific documentation is required where collection leads the CIA to acquire so much information that it cannot be quickly or effectively subjected to individualized review.[125]

Even these special rules related to unevaluated information or bulk collection, however, would not seem to preclude the kind of data aggregation, analysis, and sharing we have been discussing in this paper, at least with respect to PAI. The provisions on unevaluated information and bulk collection are clearly intended to prevent undue possession of USPII.[126] Yet precisely because PAI is by definition publicly available—and because, as we have seen, the CIA may both collect and retain PAI without special restriction even if that information contains USPII[127]—datasets known in advance to consist solely of PAI (e.g., datasets that are publicly available) would seem incapable of presenting any privacy problem of the sort these provisions were designed to prevent. To be sure, the CIA procedures do not expressly exempt "pure PAI" datasets from the rules regarding information collected in bulk and/or that has not yet been subjected to individualized review for USPII, but such a conclusion clearly follows from the language and structure of the procedures.

In sum, collection, analysis, and sharing of PAI—including nontraditional data of the sort we have been discussing—would seem permissible under the CIA procedures. What is less clear, however, is the degree to which permitted activity can scale and to what range of activities such methods could be applied.

One challenge, for instance, is the intelligence oversight system's implied assumption that U.S. person collection will generally occur in the form of incidental collection—that is, information that ends up in IC hands inadvertently, as the result of collecting against non-U.S. persons. As described above, the *incidental collection* of PAI about U.S. persons seems fairly unproblematic.

Yet specific queries about a U.S. person—such as one might wish to make if trying to determine whether a U.S. company has an adversary foreign entity in its supply chain or other commercial relationships—are more difficult. The intelligence oversight system is quite uneasy with such queries unless there is a clear foreign nexus or the query is undertaken for particular purposes such as counterintelligence. (The CIA procedures, for instance, generally only allow querying unevaluated information in circumstances such as when the query is "not designed to retrieve information concerning a U.S. person."[128]) This could make PAI analytics drawing information from federated networks of commercial data providers on a query-by-query basis—the type that are arguably best suited to permitting analysis at scale while protecting privacy—difficult where USPI is concerned, thus limiting the IC's ability to take advantage of PAI sources.[129]

DoD Intelligence Components

The procedures adopted by the Department of Defense to regulate the handling of information concerning U.S. persons are similar to those we have seen from the CIA, though in DoD parlance the focus of concern is USPI rather than USPII.[130] According to the DoD Manual on intelligence oversight, USPI is information that is

> "reasonably likely to identify one or more specific U.S. persons. USPI may be either a single item of information or information that, when combined with other information, is reasonably likely to identify one or more specific U.S. persons. Determining whether information is reasonably likely to identify one or more specific U.S. persons in a particular context may require a case-by-case assessment by a trained intelligence professional. USPI is not limited to any single category of information or technology. Depending on the context, examples of USPI may include: names or unique titles; government-associated personal or corporate identification numbers; unique biometric records; financial information; and street address, telephone number, and Internet Protocol address information."[131]

The Defense Department's definition of a "U.S. person" tracks that in E.O. 12333: a U.S. citizen or permanent resident alien, "[a]n unincorporated association substantially composed of U.S. citizens or permanent resident aliens," or a corporation incorporated in

the United States not directed and controlled by a foreign government. The DoD manual also provides a presumption rule, such that a person or organization outside the United States is presumed not to be a U.S. person—and one inside the country is presumed to be a U.S. person—until "specific information to the contrary is obtained."[132]

IC elements that are part of the DoD may intentionally collect USPI "if the information sought is reasonably believed to be necessary for the performance of an authorized intelligence mission or function assigned to the Component" and if that USPI "is publicly available."[133] PAI, in turn, is defined as information

> "that has been published or broadcast for public consumption, is available on request to the public, is accessible on-line or otherwise to the public, is available to the public by subscription or purchase, could be seen or heard by any casual observer, is made available at a meeting open to the public, or is obtained by visiting any place or attending any event that is open to the public. Publicly available information includes information generally available to persons in a military community even though the military community is not open to the civilian general public."[134]

There are some fairly minor restrictions on the handling of USPI. Personnel disseminating it, for instance, must receive training on how to disseminate such information in conformity to DoD rules,[135] and "to the extent practicable," DoD officials must "collect no more [such] information than is reasonably necessary."[136] Also "[t]o the extent practicable," moreover, "a Defense Intelligence Component should not include USPI in a dissemination … if the pertinent information can be conveyed in an understandable way without including the identifying information."[137] It is also the case that dissemination of "large amounts of unevaluated USPI" requires special approval.[138] Furthermore, it is generally required to specially mark and tag files and documents that are "reasonably believed or known to contain USPI," though this is a somewhat flexible requirement that only applies "[w]hen appropriate and reasonably possible." (In particular, "[i]n the case of certain electronic databases, if it is not reasonably possible to mark individual files containing USPI, Components may use a banner informing users before access that they may encounter USPI.")[139]

On the whole, however, the DoD system is in these regards fairly permissive. The relevant manual provides not merely that Defense Intelligence Components may collect PAI even if it contains USPI,[140] even about U.S. persons in the *United States*,[141] but also that they may *disseminate* USPI to anyone where "[t]he dissemination is to any person or entity and the information is publicly available."[142] Indeed, it also explicitly states that USPI may be disseminated—apparently whether or not it is publicly available—to "an element of DoD (including a DoD contractor)," to "any other part of the Federal Government," or to "a State, local, tribal, or territorial government," as long as the recipient is "reasonably believed to have a need to receive such information for the performance of its lawful missions or functions."[143] USPI may also be retained permanently where it was publicly available and was "lawfully collected by the Component or disseminated to the Component by another Component or element of the Intelligence Community."[144] There should, therefore, be little obstacle to DoD components of the U.S. IC acquiring, aggregating, analyzing, and sharing nontraditional data of the sort we have been discussing, provided this information meets the definition of PAI provided in the manual.

What Can Be Done?

Some Options

From the foregoing discussion, it should be clear that even under today's current arrangement of inconsistent, overlapping frameworks for regulating how government

agencies must handle PAI, there is considerable scope for accessing, analyzing, and sharing nontraditional data in ways that have by now become commonplace in many other modern contexts. As long as the information in question meets the definition of what it means to be publicly available—as indeed is much or most of the extraordinary range of data sources now available from aggregators of the digital exhaust of the modern information economy—federal officials (or contractors supporting their government missions) can do a good deal of such work without running afoul of the Privacy Act or intelligence oversight rules. Thanks to a restrictive DoD directive, the scope for such work is more limited for Defense Department components outside the U.S. IC, but—even without, for instance, some adjustment of Section 4.1 of DoDD 5200.27 to provide more clarity about the scope of authorized DoD missions related to the "protection of DoD functions" (e.g., to specify that the Department has a role in helping protect the whole military-relevant U.S. innovation base against adversary infiltration, manipulation, and strategic dependency)—they still appear to have room to do some useful nontraditional data aggregation and analysis at scale in support of missions such as DIB maintenance and supply chain security.

But given the extent to which the greatest national challenges the United States faces—among them climate change and strategic competition with China—involve cross-cutting substantive questions and policy issues in which a kaleidoscope of stakeholders must both understand their environment and act together effectively in support of some shared vision of American success, the current governance model may not be enough. Big Data analytics are now commonplace in much of the rest of modern society and are even becoming so in the information warfare practices of U.S. adversaries. In this new context, it is reasonable to ask whether we can do more to facilitate our country's ability to take advantage of the power of data in meeting America's greatest challenges. The United States would seem to have at least three main alternatives in this respect:

1. **Do nothing.**

   One possible answer, of course, might be that we should *not* do more. As noted, merely sticking with the *status quo* would not preclude all efforts to undertake data analytics to support policymaking in the face of America's wicked problems. It would certainly be possible to continue soldiering gamely along, as is done today, in employing improvised, bespoke, partial solutions on a user-by-user or mission-by-mission basis within today's legal and regulatory patchwork of rule sets. This would be inefficient, however, as well as costly in the aggregate. While still far better than shunning all data-analytic solutions entirely, however, this would provide the least value in support of genuinely WON solutions of the various alternatives we will discuss below. We fear that such a *status quo* response would not be enough.

2. **Build a "Swiss Army knife" system.**

   One possible improvement—though it would be a Herculean undertaking in various respects—might be to build a data-analytical center, perhaps managed by a nonprofit corporation or consortium acting as a trusted broker and intermediary between public and private sector stakeholders, to aggregate publicly available nontraditional data and use sophisticated analytical techniques on this information at scale (e.g., customized AI algorithms, natural language processing, and state-of-the art automation tools) *on behalf* of various diverse participants in any given critical WON effort, while adjusting this work to conform to the rule sets that currently govern data work on behalf of each user and use case.

   Such a "Swiss Army knife"-type system—the development of which the MITRE Corporation is currently exploring—would provide decision-support analytics for a wide range of participating WON stakeholders. The operational costs of doing such work—which would include not just paying for access to commercial data sources but also the expenses of managing the data and doing the analysis—would

be met on an at-cost, fee-for-service or subscription basis. (Such an approach would also entail up-front financial costs, of course, in order to meet the physical capital, computational, organizational expenses of setting up the nonprofit data center, allowing the costs passed along to stakeholders to be kept to a minimum and thus maximizing the ease and likelihood of widespread participation.)[145]

Essential to the success of such a system in supporting multiple stakeholders within today's overlapping legal and regulatory frameworks, however, would be building an extremely sophisticated management architecture that would allow this data-analytical center to navigate through all existing governance frameworks on the fly, providing services to participating stakeholders on a user-by-user and use-by-use basis. (Without some reform of the overlapping structure of regulations governing data usage, after all, any given institutional or organizational customer in the federal system—to say nothing of non-governmental stakeholders—might be operating under a somewhat different set of rules from the next one. Even for a single customer, moreover, the rules might also vary depending upon the purpose for which the information is being sought.) Because different and at least partially overlapping rules would still exist, the system would have to know how to apply the right rule in each case regarding such things as: what sorts of data can be drawn upon in performing analysis; how, by whom, to what extent, and in what form information can be handled and stored; how and to whom information could be disseminated; and requirements for record-keeping, institutional oversight, and accountability.

To make a "Swiss Army knife" service-providing data access center work, therefore, it would be necessary to devise an architecture that ascertains and closely tracks participating stakeholder data permissions, applying some sort of sophisticated gateway mechanisms within its semi- or fully-automated analytical processes to ensure each user only gets access in each case to what it is allowed to have, under its particular rules, for that particular purpose. Oversight mechanisms, moreover, would have to be built that are capable of providing at least *post hoc* insight into all of this informational juggling, so the system could assure accountability in the event of any problem. New means of pooling resources from different institutional public and private funding streams might also need to be devised to allow multiple federal sponsors and other stakeholders to cooperate in *paying* for all of this without running afoul of the various rules governing how federal funds may be spent.[146]

Such organizational and architectural sophistication is surely not impossible, but it would nonetheless be quite challenging. The difficulty and complexity of such a mechanism, moreover, would certainly reduce the efficiency and mutual situational awareness gains that might otherwise be had from seamless data-integration, analysis, and sharing at scale across a range of participating public- and private-sector stakeholders.

3. **Reform the governance system**.

   A better answer would be to reform the current federal system to harmonize, rationalize, and streamline its rules. As we emphasize further below, this need not necessarily entail any relaxation of privacy protections. Our point is merely that whatever the degree of privacy protection America's leaders choose to adopt for large-scale data analytics to help meet WON challenges, it would surely be better—if we really want to take advantage of the opportunities such analytics may provide—for there *not* to be multiple, inconsistent, and overlapping frameworks governing data usage.

   Ideally, a reform effort would harmonize data rules across the entire federal system, so that institutional leaders, data aggregators, AI developers, auditing and oversight officials, and those charged with devising

innovative use cases in support of WON missions could all work together to plan, build, and operate within a common set of well-understood parameters. A federal "backstop" of liability protections or indemnity provisions would also be helpful, to encourage private-sector participation in such a project and to protect all concerned in the face of potential litigation risk (e.g., if a litigious company is excluded from the DIB supply chain on the basis of data-driven insights into its too-close relationship with adversary entities, or if analytically-derived probabilistic conclusions the government shares with private industry ultimately turn out to be incorrect). Either way, the objective would be to create a *single* set of rules—carefully crafted to strike an appropriate balance between data-aggregative efficacy, citizen privacy, and systemic accountability—that would apply to all, ending the current "patchwork quilt" approach to data protection.

Short of such full-spectrum reform of the U.S. data-sharing system, a partial reform effort might aim to create cross-cutting rule sets on an *issue-by-issue* basis. This more incremental approach could start with creating a set of uniform standards applicable to all stakeholders in connection specifically with efforts undertaken to support one particular high-priority WON mission. Varying rules would still continue to apply when it came to *other* substantive areas of policy development and implementation, but all players would be able to work under a clear, uniform, cross-jurisdictional standard when it came to pursuing that *particular national priority mission*. (A conceptual precedent here might be DoD Directive 5200.27—which, as we have seen, is much more permissive for some investigative purposes than for others.)

Another advantage of such an approach would be that the targeted governance reforms it would entail could address the "specific query" challenge described in our discussion of intelligence oversight rules, by clarifying that U.S.-person-specific queries are not prohibited for the specific mission area in question. On its face, this might seem like it would represent some kind of derogation from strong privacy standards, but in practice it would likely result in *less*-intrusive analytics. One approach to doing sophisticated data analytics on PAI at scale is to employ bulk collection and storage—that is, to *stockpile* enormous quantities of data and then run analytical tools against the resulting "data lake." This, however, has the side effect of creating enormous reservoirs of unevaluated data in federal government hands, which, in the PAI context, likely contains a great deal of information about U.S. persons. A more privacy-protective solution would be to use tailored Application Program Interfaces (APIs) and AI-facilitated automation to allow analytics at scale primarily on the basis of specific, mission-driven queries made to commercially provided databases. With such API-based, query-specific, query-driven work, the federal system only has to manage what comes back from such queries rather than having to store the "data ocean" of privately aggregated PAI. Such query-based approaches are already straightforward with regard to non-U.S. persons. Perhaps ironically, however, clarifying their availability for U.S.-person queries might well result in vastly less information about U.S. persons being stored on government systems.

Incremental reform could begin, therefore, with a kind of pilot program. The effort would start by identifying a particular high-priority WON mission, for the specific accomplishment of which—and, initially, *only* for the accomplishment of which—data-use rules would be harmonized across the federal system.[147] To the degree this exploratory effort seemed to work, the reform project could be expanded to additional areas. The idea would be eventually to reach full, system-wide harmonization, but the reform program would proceed step by step, with all participants learning and adjusting approaches as we progress.

Supply Chain Pilot Program

Of these options, we recommend that U.S. leaders bring about governance reform. Specifically, we recommend the pilot-program approach described above. A promising place to begin with such a pilot effort would be a program to secure America's supply chains against penetration and potential manipulation by foreign adversaries. This is both an issue of pressing national security concern and an area inherently likely to involve considerable quantities of information about U.S. persons. (Indeed, to the degree the U.S. government is successful in keeping malicious foreign actors out of U.S. supply chains, it would in a sense be one objective of a national supply chain security program to *increase* the proportion of U.S. persons in such supply chains.) For this reason, institutional reform in data sharing would provide special benefits here, making this an ideal place to begin in developing better national answers to such challenges through innovations in how the government aggregates and shares PAI across a range of traditional institutional and bureaucratic stovepipes.

To this end, we recommend the creation of specific legislative authority for federal authorities to gather, analyze, and share information for purposes of understanding U.S. supply chains, identifying national security risks therein, and excluding malicious actors therefrom. Such a statute might, for instance, provide that notwithstanding any other provision of law, it is lawful to do such work whether or not the data involved contains information that would count as USPI or USPII under intelligence oversight rules, information about U.S. individuals under the Privacy Act, or other specific statutes or regulations governing the use of information about Americans. Such a new rule might even expressly authorize—and indeed *direct*, pursuant to government-wide procedures that the new statute would require be established by a date certain—all relevant departments and agencies to cooperate in sharing information for purposes of identifying and mitigating supply chain risks.[148]

Naturally, such a new law should not create an unaccountable authority to mine and disseminate USPI. Analogous to how the Attorney General is required to approve agency-specific intelligence oversight guidelines on the use of USPI, this reform might, for instance, require that the Attorney General approve the procedures federal agencies adopt for gathering, analyzing, and disseminating data containing information about U.S. persons in supply chain security programs.

The new provision, moreover, would be only *purpose-specific*, with special permission perhaps being required for information generated by such supply chain security efforts to be passed to federal officials for *other* purposes (e.g., for law enforcement or intelligence work). Moreover, analogous to how the Defense Department's new Publicly Available Information Advisory Council advises the Secretary of Defense on policy issues related to PAI,[149] the operations of such supply chain programs might also be generally overseen by a Supply Chain Risk Advisory Council. And, of course, these new rules would prohibit federal officials involved in supply chain risk mitigation of this sort from acquiring, storing, or disseminating any information about any U.S. person solely on the basis of that person's exercise of rights protected by the U.S. Constitution.

Develop Code of Ethical Principles

Additionally, we recommend the development of a "Code of Ethical Principles for the Use of Publicly Available Information" that would provide clear articulation of best practices for the ethical use of PAI. Such a code, for instance, would be analogous to the "Ethical Principles for Artificial Intelligence" adopted by the DoD for the use of AI in Pentagon programs in February 2020,[150] or the "Principles of Artificial Intelligence Ethics" promulgated by the Office of the Director of National Intelligence in June of that year.[151]

Such a code for using PAI should be developed in close consultation both with ethicists and with technical

experts intimately familiar with the range of data sources, analytical techniques, and possible applications that exist in this arena, and it should also be periodically *re*-examined to ensure its continued integrity. We do not presume to offer any definitive account here of what such a code should contain, but to provide "food for thought" that can perhaps get such discussions started, we offer the following *hypothetical* possibility:

(HYPOTHETICAL) CODE OF ETHICAL PRINCIPLES FOR THE USE OF PUBLICLY AVAILABLE INFORMATION

Preface

We recognize that sophisticated data analysis has the potential to inform policymaking and decision-making in unprecedented ways across government and in the private sector, and that the insights and situational awareness that such analytics can provide are likely to be of enormous importance in helping societies meet a range of broad, systemic challenges at the regional, national, and global levels. Given the great range of data sources that are now available either entirely openly or on a commercial basis from data providers, it is clear that publicly available information (PAI) can and should be an important source of such insight and awareness.

At the same time, we prize the values of individual privacy and autonomy, which are among the foundations of a free and democratic society and conditions for human flourishing. In this context, we recognize that—especially in the digital age—collecting and analyzing certain types or volumes of PAI may be felt to raise privacy concerns, at least where such information tends to identify specific individuals and to provide a third party with significant visibility into aspects of those individuals' lives that they consider to be private matters. If gathered and handled without appropriate controls, PAI has the potential to cause embarrassment, inconvenience, or even harm. (We also recognize that restrictions upon the use of information that is in fact available publicly could itself raise concerns about infringements of free speech and public expression.)

Any approach to managing the use of PAI must thus be informed both by the potential value of sophisticated modern data analytics and by the possibility of undue invasiveness, or even outright misuse. Ethical users of PAI should be aware of this possible tension, should acknowledge and seek to manage it, and should be open and honest about how they handle these challenges. PAI use should follow a clearly articulated understanding of ethical best practices.

Principles

To this end, we commit to acquiring, holding, aggregating, analyzing, and sharing PAI according to the following principles:

1. PAI should not be gathered, held, aggregated, analyzed, or shared for anything other than legitimate, lawful, and authorized purposes, and any user should be able (and, if challenged, expected) to articulate and explain the propriety of those purposes.
2. No PAI should be acquired, held, or shared beyond the minimum amount or degree necessary to effectively perform the specific function or functions for which it is gathered.
3. Each type or category of PAI data should be carefully vetted for its relative quality, integrity, and reliability, and for its effectiveness in contributing to mission accomplishment, before such type or category is incorporated into ongoing data acquisition, analytical, and information-sharing processes. Individual vendors, as applicable, should also be carefully vetted to ensure their own quality, integrity, and reliability, before being engaged to provide data.
4. Privacy-protecting mechanisms for handling data (e.g., appropriate anonymization, data filters, access permissions, retention timelines) should be developed, adopted, and employed to the greatest extent possible consistent with the effective accomplishment of such functions.

(HYPOTHETICAL) CODE OF ETHICAL PRINCIPLES FOR THE USE OF PUBLICLY AVAILABLE INFORMATION
(CONTINUED)

5. Where particular types or volumes of data may tend to offer significant insight into aspects of the lives of individual persons that these persons are highly likely to consider private matters, additional controls should be employed on a case-by-case basis as warranted by the circumstances.
6. All acquisition, retention, analysis, and sharing of PAI and PAI-based analytical conclusions should be carried out using means of communication, cybersecurity methods, and personnel vetting procedures appropriate to the degree of sensitivity involved.
7. Effective policies should be put in place to review, on an ongoing basis, whether these ethical principles are being followed and whether the conclusions that underlie all these abovementioned decisions remain valid.
8. Ethical users of PAI should retain a record of all significant decisions made under these principles (e.g., in evaluating data fit to mission, vetting data types and sources, or determining appropriate protections to apply), and should be able to articulate and explain these decisions if and when asked. All such determinations should be auditable and explicable.
9. Ethical users should not acquire, store, or disseminate any PAI about any U.S. person solely on the basis of that person's exercise of rights protected by the U.S. Constitution.
10. Government users of PAI-based analytics, as well as any contractors employed by them to undertake such work, should establish and maintain and appropriate processes for (a) overseeing and auditing the use of PAI, (b) making available functions analogous to those of a privacy ombudsman and/or whistleblower protections to help ensure the ongoing integrity of PAI use, and (c) appropriately recording and reporting on the existence and function of such procedures and the results of any such audits (e.g., to Congress or to the public, as applicable

Conclusion

It is essential that we take full advantage of what sophisticated analysis of publicly available nontraditional data has to offer to America's leaders in helping meet our country's most pressing WON challenges. This, however, will require access to vast amounts and breadths of data sources—not to mention state-of-the-art analytical tools, automated data management, and bespoke AI algorithms—only now becoming available. As the term "whole of nation" implies, moreover, succeeding in these endeavors will require us to access, understand, and share information across traditional institutional boundaries in unprecedented ways. Yet the U.S. system is not well organized to do this, and real success will likely require the reform and rationalization of existing data governance frameworks.

It is worth stressing, in conclusion, that success in innovative leveraging of publicly available nontraditional data in support of WON strategy need *not* require compromising American values. The authors of this paper, in fact, share the widespread desire to avoid having modern data analytics run roughshod over the privacy of American citizens.

We do not take a position here on exactly *how much* protection should be built into a future data governance framework to help our nation meet the pressing need to respond to WON challenges. Various positions are certainly conceivable—any of which would represent,

in some sense, a legitimate choice if undertaken with full awareness of the potential consequences. Given that Google, Facebook, Twitter, and Wall Street hedge funds are already undertaking such analysis, who do the American people trust in this respect? Should it be permissible for our own elected and constitutionally and statutorily constrained government to analyze the publicly available digital exhaust of today's economy for the public good—and if so, what procedures are most appropriate in this regard? Or should we attempt to prohibit *all* such work, even though entities such as China's security services and Russia's Internet Research Agency may be little inclined to heed such restrictions?

For our part, we hope that this country will be able to do more to take advantage of what data analytics has to offer, and in a way that remains consistent with American values. Moreover, to our eye, this is very possible.

But whatever the standard of protection and accountability that is adopted, we strongly feel more should be done to make it a *uniform* standard. Over time, it is surely a recipe for willful inadequacy and failure to subject our country's response to WON problems to a data-management architecture that regulates the use of PAI data on a "patchwork" basis through multiple, inconsistent, and overlapping frameworks. This is not to fault today's existing frameworks or to blame their authors. Those frameworks evolved at different times and to serve different purposes, and they developed before it had become clear what a remarkable range of datasets are now available, what a powerful suite of analytical tools exist with which to digest digital exhaust, and the many uses to which such data are routinely being put in every advanced sector of the modern information economy e*xcept* government. The drafters of those rules did not have the benefit of this knowledge, and there is no reason to fault them for having woven the inconsistent patchwork quilt under which we now labor. Nevertheless, *we do* now know what is possible, and the authors of this paper believe America can now find a better way for the future.

We suggest a modest approach to harmonizing standards for the use of PAI, one that could begin with a pilot program that would seek to begin down a path to better answers by starting with a program to help secure America's critical supply chains against foreign-adversary control or manipulation. What's more, we should help reassure all stakeholders that Americans' privacy interests are being protected and that PAI is always being used in appropriate and accountable ways, so that we both protect privacy *and* are able to take advantage of the significant benefits available to our nation from cutting-edge data analytics in addressing national challenges. To that end, we urge that steps be taken to develop, clearly articulate, and adhere scrupulously to a new "Code of Ethical Principles for the Use of Publicly Available Information," on which we also have offered at least some tentative thoughts.

To meet our nation's needs by equipping the United States to meet the WON challenges it faces, the White House, Congress, and indeed all public and private sector stakeholders should turn their attention to these questions. It is time to reform the federal data-management architecture in ways that facilitate WON competitive success while protecting the values the American people cherish most.

About the Authors

Christopher Ford is a MITRE Fellow and Director of MITRE's Center for Strategic Competition, as well as a Visiting Fellow at Stanford University's Hoover Institution. Until January 2021, he served as U.S. Assistant Secretary of State for International Security and Nonproliferation, also fulfilling the duties of the Under Secretary for Arms Control and International Security. Before his most recent service at the Department of State, Ford served as Special Assistant to the President and Senior Director for WMD and Counterproliferation at the National Security Council. He is the author of The Mind of Empire: China's History and Modern Foreign Relations (2010) and China Looks at the West: Identity, Global Ambitions, and the Future of Sino-American Relations (2015).

Marin Halper is MITRE's Vice President for Cross-Cutting Priorities, addressing complex system-of-system challenges that bridge across national security sponsors, including the Department of Defense, the military services, and the intelligence community—and increasingly require integrated solutions that span national security, other public sector government departments and agencies, and private industry. Marin previously co-founded and led MITRE's Cross-Cutting Urgent Innovation Cell (CUIC) and Non-Traditional Data Employment for Competition Priority.

Andrea McFeely is a Principal Analytic Systems Engineer with MITRE's CUIC and co-leads the MITRE NSEC Non-Traditional Data Employment for Competition Priority. She also currently leads a three-year research project focused on bridging innovation gaps between commercial companies and the U.S. government to foster more-effective strategic competition awareness. She is a former U.S. government intelligence analyst and diplomat.

The views expressed herein are the authors' own, and do not necessarily reflect those of anyone else.

Endnotes

1 "Biden Administration Roadmap to Build an Economy Resilient to Climate Change Impacts," White House Fact Sheet (October 15, 2021), *available at* https://www.whitehouse.gov/briefing-room/statements-releases/2021/10/15/fact-sheet-biden-administration-roadmap-to-build-an-economy-resilient-to-climate-change-impacts/.

2 "Biden Administration and Private Sector Leaders Announce Ambitious Initiatives to Bolster the Nation's Cybersecurity," White House fact sheet (August 25, 2021), *available at* https://www.whitehouse.gov/briefing-room/statements-releases/2021/08/25/fact-sheet-biden-administration-and-private-sector-leaders-announce-ambitious-initiatives-to-bolster-the-nations-cybersecurity/.

3 Taite R. McDonald, Kevin L. Turner, Eric S. Crusius, Andrew K. McAllister, & Hannah M. Coulter, "Biden Administration Report Outlines Strategy to Invigorate Domestic Supply Chains," Holland & Knight Alert blog (June 11, 2021), *available at* https://www.hklaw.com/en/insights/publications/2021/06/biden-administration-report-outlines-strategy-to-invigorate-domestic.

4 *See* Executive Order 14017, "Executive Order on America's Supply Chains" (February 24, 2021), *available at* https://www.whitehouse.gov/briefing-room/presidential-actions/2021/02/24/executive-order-on-americas-supply-chains/.

5 *See, e.g.,* Tom Christensen & Per Laegreid, "The Whole-of-Government Approach to Public Sector Reform," *Public Administration Review,* vol. 67, no. 6 (November-December 2007), at 1059, 1063.

6 Peter Williams, Jan Gravesen, & Trinette Brownhill, "Achieving Joined-Up Government," *Government Executive* (November 24, 2015), *available at* https://www.govexec.com/management/2015/11/achieving-joined-government/123952/; *see* also, generally, Duke K. McCall, Douglas A. Hastings, & Meredith L. Compton, "Biden-Harris Administration's 'All of Government' Approach to Addressing Climate Change and Environmental Justice," *Morgan Lewis* blog (April 22, 2021), *available at* https://www.morganlewis.com/pubs/2021/04/biden-harris-administrations-all-of-government-approach-to-addressing-climate-change-and-environmental-justice; Christopher Hood, "The Idea of Joined-Up Government: A Historical Perspective," in *Joined-Up Government* (Vernon Bogdanor, ed.) (Oxford: Oxford University Press, 2005), at 22.

7 The term "wicked problems" comes from the German mathematician and designer Horst Rittel, who coined it in the 1960s. As recounted by Richard Buchana, Rittel "sought an alternative to the linear, step-by-step model of the design process being explored by many designers and design theorists." For Rittel, in his own words, wicked problems are "a class of social system problems which are ill-formulated, where the information is confusing, where there are many clients and decision makers with conflicting values, and where the ramifications in the whole system are thoroughly confusing." Richard Buchanan, "Wicked Problems in Design Thinking," *Design Issues,* vol. 8, no. 2 (Spring 1992), at 5, 15, *available at* https://web.mit.edu/jrankin/www/engin_as_lib_art/Design_thinking.pdf.

8 Christensen & Laegreid, *supra*, at 1059-61.

9 Mohamed Hairul Othman & Rozilawati Razali, "Whole of Government Critical Success Factors towards Integrated E-Government Services: A Preliminary Review," *Jurnal Pengurusan,* vol. 53 (October 2018), at 73-82, *available at* https://www.researchgate.net/publication/335758341_Whole_of_Government_Critical_Success_Factors_towards_Integrated_E-Government_Services_A_Preliminary_Review.

10 *See, e.g.,* CCP, Central Committee and the State Council, *Outline of the National Innovation-Driven Development Strategy* (published by Xinhua News Agency) (May 19, 2016), *available at* https://cset.georgetown.edu/publication/outline-of-the-national-innovation-driven-development-strategy/ [Chinese language source: http://www.xinhuanet.com/politics/2016-05/19/c_1118898033.htm]; National People's Congress, *Outline of the People's Republic of China 14th Five-Year Plan for National Economic and Social Development and Long-Range Objectives for 2035* (published by Xinhua News Agency) (March 12, 2021), *available at* https://cset.georgetown.edu/wp-content/uploads/t0284_14th_Five_Year_Plan_EN.pdf [Chinese language source https://perma.cc/73AK-BUW2].

[11] Christopher Ford, "Charting New Horizons: Technology and U.S. Competitive Success," *MITRE Corporation* (August 2021), at 4, *available at* https://www.mitre.org/sites/default/files/publications/pr-21-2393-charting-new-horizons-technology-and-u.s.-competitive-success.pdf.

[12] Christopher Hood, "The Idea of Joined-Up Government: A Historical Perspective," in *Joined-Up Government* (Vernon Bogdanor, ed.) (Oxford: Oxford University Press, 2005), at 23.

[13] Ford, "Charting New Horizons," *supra,* at 4.

[14] Christensen & Laegreid, *supra,* at 1063

[15] *Id.* at 1062 (internal citations omitted).

[16] Williams, Gravesen, & Brownhill, *supra; see also, e.g.,* Othman & Razali, *supra.*

[17] Executive Order 14017, *supra.*

[18] Jere Kolehmainen, "Benefits of Data Aggregation in Supply Chains," *Youredi* blog (September 22, 2020), *available at* https://www.youredi.com/blog/benefits-of-data-aggregation-in-supply-chains.

[19] Tucker Bailey, Edward Barriball, Arnav Daly, & Ali Sankur, "A practical approach to supply-chain risk management," *McKinsey* blog (March 8, 2019), *available at* https://www.mckinsey.com/business-functions/operations/our-insights/a-practical-approach-to-supply-chain-risk-management.

[20] "Supply Chain Mapping Combined with Risk Assessment," *SupplyChainBrain* blog (March 25, 2014), *available at* https://www.supplychainbrain.com/articles/18577-supply-chain-mapping-combined-with-risk-assessment.

[21] Kolehmainen, *supra.*

[22] Bailey, et al., *supra.*

[23] *Cf.,* Scott E. Page, *Diversity and Complexity* (Princeton, NJ: Princeton University Press, 2011), at 6-7 & 17 (discussing elements that give rise to complexity). Note that in this context, "complex" is a technical term the meaning of which is different from how the word is used colloquially, which equates merely to something like "very complicated."

[24] A number of scholars influenced by modern Complexity Science have noted this problem. *See generally, e.g.,* Göktug Morçöl, *A Complexity Theory for Public Policy* (New York: Routledge, 2012), at 9, 34, 86, 89, & 143-45; David Colander & Roland Kupers, *Complexity and the Art of Public Policy: Solving Society's Problems from the Bottom Up* (Princeton, NJ: Princeton University Press, 2014), at 5-8; Christopher Ford, "Policymaking at the Edge of Chaos: Musings on Political Ideology Through the Lens of Complexity," Hudson Institute (January 2011), at 7-8, *available at* https://www.hudson.org/content/researchattachments/attachment/857/conceptualizing_ideology.pdf; Organization for Economic Co-operation and Development Global Science Forum, *Applications of Complexity Science for Public Policy: New Tools for Finding Unanticipated Consequences and Unrealized Opportunities* (September 2009), at 2, *available at* https://paperzz.com/doc/9339201/applications-of-complexity-science-for-public-policy; Buchanan, *supra*, at 15.

[25] Cf., Larry English, "Digital Exhaust: The Most Valuable Asset Your Organization Owns, But Isn't Using," *Forbes* (February 1, 2021), *available at* https://www.forbes.com/sites/larryenglish/2021/02/01/digital-exhaust-the-most-valuable-asset-your-organization-owns-but-isnt-using/?sh=7bdc843332ac. English is referring to the electronic data created inside a particular company by its own routine activities, but the term might as well be applied to the electronic "noise" of transactional, commercial, and communicative engagement between corporations, individuals, and other entities throughout the modern digital economy—which, as we will *see*, is these days increasingly being collected and commoditized by private sector data aggregators.

[26] As we use this informal term here, it loosely describes data that earlier generations generally neither created nor that analysts were able to acquire and aggregate it at significant scale.

[27] AlternativeData.org website, *available at* https://alternativedata.org/alternative-data/ (visited February 28, 2022).

[28] Alicia McElhaney, "Big Data: Too Popular for its Own Good?" *Institutional Investor* (August 9, 2018), *available at* https://www.institutionalinvestor.com/article/b19fsq17p6zp5n/Big-Data-Too-Popular-for-its-Own-Good.

[29] Nicholas Samuel, "7 Best Data Aggregation Companies," *Hevo* blog (February 5, 2021), *available at* https://hevodata.com/learn/7-best-data-aggregation-companies/.

[30] Any Patrizio, "What is Data Aggregation?" *Datamation* blog (May 7, 2021), *available at* https://www.datamation.com/big-data/data-aggregation/.

[31] Janna Anderson & Lee Rainie, "Main Findings: Influence of Big Data in 2020," Pew Research Center (July 20, 2012) (quoting Bryan Trogdon), *available at* https://www.pewresearch.org/internet/2012/07/20/main-findings-influence-of-big-data-in-2020/.

[32] Such concerns were raised, for instance, by some of the more pessimistic respondents in one 2012 survey of data experts. *See* Anderson & Rainie, *supra*.

[33] *See, e.g.,* William Safire, "You Are a Suspect," *New York Times* (November 14, 2002), *available at* https://www.nytimes.com/2002/11/14/opinion/you-are-a-suspect.html.

[34] *See, e.g.,* Anderson & Rainie, *supra*.

[35] AI, for instance, may allow much higher volumes of data to be analyzed and may allow users to discern "patterns in those data that would, for reasons of both volume and dimensionality, otherwise be beyond the capacity of human interpretation." Kathleen McKendrick, "Artificial Intelligence Prediction and Counterterrorism," Chatham House (August 2019), at 2. *available at* https://www.chathamhouse.org/sites/default/files/2019-08-07-AICounterterrorism.pdf. The use of multiple data sources, moreover, can help improve overall analytical quality by reducing dependence upon any one source and allowing analysts to cross-check the quality of each source through a sort of cross-referencing against others. *See* Marybeth Regan, "Collaboration: The Power of Data Aggregation," *Heathleaders* blog (April 22, 2008), *available at* https://www.healthleadersmedia.com/innovation/collaboration-power-data-aggregation.

[36] There does not *see*m to be a canonical definition of "Big Data," but the term is generally used to refer to "datasets whose size is beyond the capability of typical database software tools to capture, store manage and analyse." McKendrick, *supra*, at 6.

[37] *See, e.g.,* Rock Content Data "40 Places to Find Open Data on the Web," Rockcontent blog (March 20, 2012), *available at* https://rockcontent.com/blog/data-sources/; *see also, e.g.,* https://freegisdata.rtwilson.com (list of openly available geographic data sets available for loading into Geographic Information System applications) (visited February 14, 2022); https://data.imf.org/?sk=388DFA60-1D26-4ADE-B505-A05A558D9A42&sId=1479329132316 (list of data sets available from the International Monetary Fund on macroeconomic and financial data) (visited February 14, 2022); https://libguides.umn.edu/HealthStatistics (list of health statistics and related data sources) (accessed February 14, 2022); and https://open-power-system-data.org/data-sources (list of data sources "that are helpful for power system modeling of Europe") (visited February 14, 2022).

[38] *See, e.g.,* https://www.mobiusservices.com/data-aggregation-as-a-service (visited March 1, 2022).

[39] *See, e.g.,* https://alternativedata.org/alternative-data/ (visited February 28, 2022) (listing illustrative providers for such datasets); McElhaney, *supra* (citing survey in 2018 by Greenwish Associates 2018 Data Customer Journey Study and providing list of data types commonly employed); James Pfeiffer, "The Basics of Financial Data Aggregation," *Terrapin Technologies* blog (December 16, 2021), *available at* https://terrapintech.com/basics-of-financial-data-aggregation/. *See also* "Hedge funds' use of alternative data tipped to surge, new industry study finds," *Hedgeweek* blog (April 2, 2020), *available at* https://www.hedgeweek.com/2020/05/04/285283/hedge-funds-use-alternative-data-tipped-surge-new-industry-study-finds.

[40] *See, e.g.,* Renee DiResta, "How the Tech Giants Created What DARPA Couldn't," *Wired* (May 29, 2018), *available at* https://www.wired.com/story/darpa-total-informatio-awareness/. Nor are such private sector uses necessarily predatory, of course, for some can be of considerable benefit to ordinary citizens. Credit card companies protect their customers by spotting fraud through algorithms that flag "unusual activity" in accounts, for instance, and service providers such as Netflix and Amazon make recommendations to their customers based upon data analytics that are often quite sophisticated.

Google search trends, moreover, have been used to identify public health problems well before traditional data is available from healthcare providers. *See* Anderson & Rainie, *supra*. (By at least one account, moreover, some consumers may also opt-in to "open banking" mechanisms wherein their financial data is shared with third-party financial service providers to permit more "personalized" service. *See* Darcy Tyrrell, "How Financial Aggregators Benefit consumers and Businesses," *Yodlee* blog (July 29, 2020), *available at* https://www.yodlee.com/data-aggregation/financial-aggregators.)

[41] *See* John Raidt, "7 Great Ways That Data Can Benefit Society," U.S. Chamber of Commerce Foundation (May 23, 2016), *available at* https://www.uschamberfoundation.org/blog/post/7-great-ways-data-can-benefit-society-0.

[42] Marybeth Regan, "Collaboration: The Power of Data Aggregation," *Heathleaders* blog (April 22, 2008), *available at* https://www.healthleadersmedia.com/innovation/collaboration-power-data-aggregation.

[43] "The Importance of Data Aggregation in Healthcare," Tiga Healthcare Technologies blog (undated), *available at* https://www.tigahealth.com/the-importance-of-data-aggregation-in-healthcare/. AI algorithms fed by and trained on large healthcare datasets have already helped speed up the introduction of messenger RNA-based vaccines such as those now being used to rein in COVID-19. *See, e.g.,* United Nations Counter-Terrorism Centre & United Nations Interregional Crime and Justice Research Institute, *Countering Terrorism Online with Artificial Intelligence: An Overview for Law Enforcement and Counter-Terrorism Agencies in Southeast Asia* (2021) [hereinafter "United Nations, 'Countering Terrorism Online'"], at 12, *available at* https://www.un.org/counterterrorism/sites/www.un.org.counterterrorism/files/countering-terrorism-online-with-ai-uncct-unicri-report-web.pdf.

[44] Anderson & Rainie, *supra* (quoting then-NSF Director Subra Suresh).

[45] Travis J. Osterman, May Terry, & Robert S. Miller, "Improving Cancer Data Interoperability: The Promise of the Minimal Common Oncology Data Elements (mCODE) Initiative," *JDO Clinical Cancer Informatics* (2020), at 993.

[46] *See, e.g.,* United Nations, "Countering Terrorism Online," *supra*, at 7, 12, & 20-25; Nicole A. Softness, "Social Media and Intelligence: The Precedent and Future for Regulations," *American Intelligence Journal*, vol. 34, no. 1 (2017), at 32, 32; *see also* generally Paul K. Davis, Walter L. Perry, Ryan Andrew Brown, Douglas Yeung, Parisa Roshan, & Phoenix Voorhies, "Using Behavior Indicators to Help Detect Potential Violent Acts: A Review of the Science Base," RAND Corporation (2013) (providing detailed analysis of range of available predictive and analytic tools), *available at* https://www.rand.org/content/dam/rand/pubs/research_reports/RR200/RR215/RAND_RR215.pdf.

[47] McKendrick, *supra*, at 8; Davis et al., *supra*, at 71-74 & 143; Softness, *supra*, at 33 (citing Alexandra Mateescu, Douglas Brunton, Alex Rosenblat, Desmond Patton, Zachary Gold, & Danah Boyd, "Social Media Surveillance and Law Enforcement," *Data & Civil Rights: A New Era of Policing and Justice* [October 27, 2015]). (Such pattern analytics have also helped give rise to the phenomenon of social media intelligence, or "SOCMINT." As early as 2014, at least 80 percent of U.S. federal, state, and local law enforcement were already reported to turn regularly to social media platforms for intelligence. Softness, *supra*, at 32.)

[48] *See, e.g.,* Issie Lapowsky, "House Democrats Release 3,500 Russia-linked Facebook Ads," *Wired* (May 18, 2018) (discussing use of social-media targeting by Russia's Internet Research Agency cyber propaganda organ), *available at* https://www.wired.com/story/house-democrats-release-3500-russia-linked-facebook-ads/.

[49] Executive Order 14017, *supra*.

[50] IJavier Garcia-Bernardo, Jan Fichtner, Frank W Takes, & Eelke M. Heemskerk, "Uncovering Offshore Financial Centers: Conduits and Sinks in the Global Corporate Ownership Network," *Nature.com* website (July 24, 2017), *available at* https://www.nature.com/articles/s41598-017-06322-9.pdf. In identifying "global ownership networks," the authors treated "a series of companies [as being] connected in a chain if for each two directly subsequent entities *A* and *B*, it holds that firm *A* is owned by firm *B*, i.e., there is a link between them in the ownership network."

[51] In 2019, for instance, on study used a "dataset that codifies control for 42,700 listed firms, incorporated in 127 countries" covering roughly 90 percent world stock market capitalization to create its own map of depicting the global distribution of firms controlled by dominant shareholders or state entities. Gur Aminadav & Elias Papaioannou, "Corporate Control Across

the World," *Harvard Law School Forum on Corporate Governance* (October 10, 2019), *available at* https://corpgov.law.harvard.edu/2019/10/10/corporate-control-across-the-world/. In 2020, another group of scholars drew upon databases containing "the ownership holdings in 49 million companies worldwide by 69 million shareholders," showing through their resulting analysis that sovereign national governments—and in particular, the People's Republic of China—were much more important worldwide than previously thought, as a result of being major loci of potential corporate "network power" through latticeworks of dispersed ownership linkages. Takayuki Mizuno, Shohei Doi, & Shuhei Kurizaki, "The power of corporate control in the global ownership network," *PLoS One*, vol. 15, no. 8 (August 27, 2020), *available at* https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7451575/.

[52] *See, e.g.,* Assistant Secretary of State Christopher Ford, "Technology Transfer De-Risking: A New and Growing Need," *Arms Control and International Security Papers,* vol. 1, no. 23 (December 7, 2020), *available at* https://irp-cdn.multiscreensite.com/ce29b4c3/files/uploaded/ACIS%20Paper%2023%20-%20Tech%20Transfer%20De-Risking.pdf.

[53] Davis et al., *supra*, at 75.

[54] *Id.* at 137.

[55] *Id.* at 148.

[56] McKendrick, *supra*, at 19.

[57] *Id.* at 19 & 33.

[58] David Kris, "The CIA's New Guidelines Governing Publicly Available Information," *Lawfare* (March 21, 2017), *available at* https://www.lawfareblog.com/cias-new-guidelines-governing-publicly-available-information.

[59] I have used "acquiring" here because one needs to be somewhat careful with the word "collecting." In everyday speech it is generally a synonym for "acquiring," but as used within the U.S. national security community, "collection" may be taken to refer specifically to *intelligence* collection. *See, e.g.,* CIA, "Central Intelligence Agency Intelligence Activities: Procedures Approved by the Attorney General Pursuant to Executive Order 12333 (January 17, 2017) [hereinafter "CIA Procedures"], at § 12.3 (defining "collection" as "the receipt of information by the CIA for official purposes"), *available at* https://www.cia.gov/static/54871453e089a4bd7cb144ec615312a3/CIA-AG-Guidelines-Signed.pdf. This, as we shall see, can bring into play a somewhat different set of rules regarding the handling of information about U.S. persons.

[60] Department of Defense Directive 3115.18, *DoD Access to and Use of Publicly Available Information* (June 11, 2019) [hereinafter "DoDD 3115.18"], at § G2 (Glossary), *available at* https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/311518p.pdf?ver=2020-08-20-121154-277.

[61] Quoted in Office of the Director of National Intelligence, *Civil Liberties and Privacy Guidance for Intelligence Community Professionals: Properly Obtaining and Using Publicly Available Information* (approved for public released by DNI Pre-Pub 20140708) (July 2011) [hereinafter *"ODNI, Civil Liberties and Privacy Guidance"*], at 4, *available at* https://www.dni.gov/files/documents/CLPO/CLPO%20Publication_Publicly%20Available%20Information_July%202011%20-%20Public%20Release%20Version.pdf.

[62] ODNI, *Civil Liberties and Privacy Guidance, supra*, at 5. The focus upon information that "is accessible to the public" covers data that essentially anyone could in theory gather from accessing the internet on his or her own—whether at the "surface web" level that is indexable by normal search engines, at the "deep web" level accessible only *through* such indexable "surface" pages but that nonetheless can be visited with a standard browser, or even at the "Dark Web" level behind heavy encryption and specialized access tools such as The Onion Router (a.k.a. TOR)—as long as no special third-party vetting or other permission is required. The standard is whether or not a member of the public could freely and lawfully access the data, even if such access would require above-average computer skills. *See* Henricks, *supra*, at 22 & 28.

[63] Executive Order 12333 (December 4, 1981), at § 2.3(a), *available at* https://www.archives.gov/federal-register/codification/executive-order/12333.html; *see also* Steven C. Henricks, "Social Media, Publicly Available Information, and the Intelligence Community," *American Intelligence Journal,* vol. 34, no. 1 (2017), at 21, 28 (noting that under E.O. 12333 an IC component

can buy PAI, collect it, and receive it from another IC component). As a corollary, information is *not* PAI if it is protected by a specific statute—such as the Telecommunications Act, Electronic Communications Privacy Act, FISA, Right to Financial Privacy Act, Fair Credit Reporting Act (FCRA), Health Insurance Portability and Accountability Act (HIPAA), Family Educational Rights and Privacy Act, Driver's Privacy Act, Video Privacy Protection Act, and Cable Communication Policy Act—if it is only available to an intelligence professional by virtue of his or her position or status as a member of the IC, if "specialized [intelligence] tradecraft or skills" are needed in order to obtain it, or if it is protected by the protection the Fourth Amendment to the U.S. Constitution gives against unreasonable searches or seizures. *ODNI, Civil Liberties and Privacy Guidance, supra*, at 4-7.

64 ODNI, *Civil Liberties and Privacy Guidance, supra*, at 3.

65 United Nations, "Countering Terrorism Online," *supra*, at 23; *see also, e.g.,* Henricks, *supra*, at 24 ("Open-source intelligence … is intelligence that is produced from publicly available information and collected, exploited, and disseminated in a timely manner to an appropriate audience for the purpose of addressing a specific intelligence requirement. … The production of open-source intelligence is a valuable intelligence discipline that must be integrated into intelligence tasking, collection, processing, exploitation, and dissemination to ensure that United States policy makers are fully and completely informed.") (*quoting* 119 Stat. 3411 [2006]).

66 "Risks, benefits, and 'why there needs to be a change,'" *Federal News Network* (October 18, 2019) (quoting Sean Heritage), *available at* https://federalnewsnetwork.com/federal-insights/2019/10/publicly-available-information-risks-benefits-and-why-there-needs-to-be-a-change/.

67 DoDD 3115.18, *supra*, at ¶ 1.2(b). It also expressly instructs DoD component heads to "train personnel to DoD standards and on appropriate use of PAI" for purposes such as "force protection, warning, and other missions to protect personnel and mitigate foreign intelligence, counterintelligence, terrorist, and other threats." *Id.* at ¶ 2.10(e).

68 *Id.* at ¶ 1.2(g).

69 *Id.* at ¶ 1.3(e) (emphasis added). PAI tools are defined as "[a]pplications or capabilities that mine or derive meaning from PAI data and that acquire, analyze, store, and disseminate PAI. PAI tools and those used for intelligence missions may overlap." *Id.* at § G2 (Glossary). The Directive also established DOD PAI Advisory Council (PAC), co-chaired at the Under Secretary level within the Department, to be an advisory body that "serves as the senior DoD deliberative body to address DoD policy issues related to PAI. The PAC identifies, recommends, and promotes standard and supporting policies related to the use of PAI (e.g., regarding oversight, training, lexicon, and identity management)." *Id.* at ¶¶ 2.1(d), 2.2(d), & 3.1.

70 Henricks, *supra*, at 25 & 28; Courtney Weinbaum, "The Intelligence Community's Deadly Bias Toward Classified Sources," *RAND Corporation* blog (April 12, 2021), *available at* https://www.rand.org/blog/2021/04/the-intelligence-communitys-deadly-bias-toward-classified.html; Harry Kimsley, "In OSINT we trust?" *The Hill* (September 1, 2021), *available at* https://thehill.com/opinion/national-security/569738-in-osint-we-trust; Bob Ashley & Neil Wiley, "How the Intelligence Community Can Get Better at Open Source Intel," *Defense One* (July 16, 2021), *available at* https://www.defenseone.com/ideas/2021/07/intelligence-community-open-source/183789/.

71 ODNI, *Civil Liberties and Privacy Guidance, supra*, at 1.

72 *See* Byron Tau & Dustin Volz, "Defense Intelligence Agency Expected to Lead Military's Use of Open-Source Data," *Wall Street Journal* (December 10, 2021), *available at* https://www.wsj.com/articles/defense-intelligence-agency-expected-to-lead-militarys-use-of-open-source-data-11639142686.

73 Tau & Volz, *supra*.

74 *See, e.g.,* Christopher Ford & Charles Clancy, "A 'Horizon Strategy' Framework for Science and Technology Policy for the U.S. Innovation Economy and America's Competitive Success," *MITRE Corporation* (2021), *available at* https://www.mitre.org/sites/default/files/publications/pr-21-1440-horizon-strategy-framework-science-technology-policy.pdf.

[75] *Carpenter v. United States,* U.S. Supreme Court, no.16-1402 (June 22, 2018) slip opinion, at 10 (Chief Justice Roberts, for the Court), ("Given the unique nature of cell phone location records, the fact that the information is held by a third party does not by itself overcome the user's claim to Fourth Amendment protection. Whether the Government employs its own surveillance technology as in *Jones* or leverages the technology of a wireless carrier, we hold that an individual maintains a legitimate expectation of privacy in the record of his physical movements as captured through CSLI [cell site location information]. The location information obtained from Carpenter's wireless carriers was the product of a search."), *available at* https://www.supremecourt.gov/opinions/17pdf/16-402_h315.pdf. The holding in *Carpenter* was a considerable change from prior interpretations, in which providing data to a third party essentially waived all Fourth Amendment privacy protections vis-à-vis the government. *Smith v. Maryland,* U.S. Supreme Court, no. 78-5374 (June 20, 1979) (Justice Blackmun for the Court) ("… [W]e doubt that people in general entertain any actual expectation of privacy in the numbers they dial. All telephone users realize that they must 'convey' phone numbers to the telephone company, since it is through telephone company switching equipment that their calls are completed. … [P]etitioner can claim no legitimate expectation of privacy here. When he used his phone, petitioner voluntarily conveyed numerical information to the telephone company and 'exposed' that information to its equipment in the ordinary course of business."), *available at* https://caselaw.findlaw.com/us-supreme-court/442/735.html.

[76] Privacy Act of 1974, 5 U.S.C. §§ 552a et seq.

[77] Executive Order 12333, "United States Intelligence Activities," as amended by Executive Orders 13284 (2003), 13355 (2004) and 13470 (2008) [hereinafter "E.O. 12333"], *available at* https://dpcld.defense.gov/Portals/49/Documents/Civil/eo-12333-2008.pdf.

[78] U.S. Department of Justice, "Overview of the Privacy Act: 2020 Edition" (2020), from the Introduction, *available at* https://www.justice.gov/opcl/overview-privacy-act-1974-2020-edition/introduction#LegHistory.

[79] *Id.* (citing S. Comm. on Gov't. Operations & H.R. Comm. on Gov't. Operations, 94th Cong., Legislative History of the Privacy Act of 1974 S. 3418 (Public Law 93-579): Source Book on Privacy at 4 (Comm. Print 1976), *available at* https://www.justice.gov/opcl/paoverview_sourcebook; and *Records, Computers, and the Rights of Citizens: Report of the Secretary's Advisory Committee on Automated Personal Data Systems,* DHEW Publication No. (OS) 73-94 (July 1973), *available at* https://www.justice.gov/opcl/docs/rec-com-rights.pdf).

[80] 5 U.S.C. § 552a(a)(2).

[81] *Id.* at § 552a(a)(4).

[82] 5 U.S.C. §§ 552a(b)(1) & (3). There are other specified exceptions, but they are less relevant here. One such exception is for law enforcement activity. *See, e.g.,* Department of Defense 5400.11-R, *Department of Defense Privacy Program* (May 14, 2007) [hereinafter "DoD 5400.11-R"], at ¶ C4.2.7, *available at* https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodm/540011r.pdf; *see also, e.g.,* Henricks, *supra*, at 28 (citing *Jabara v. Webster,* 691 F.2d 272, 280 (6th Cir. 1982)).

[83] 5 U.S.C. § 552a(a)(5).

[84] *Id.* at § 552a(a)(7). With regard to "routine use," Defense Department regulations provide that "[r]ecords may be disclosed outside the Department of Defense pursuant to a routine use that has been established for the system of records that contains the records. … A routine use shall: … Be compatible with the purpose for which the record was collected; … Identify the persons or organizations to whom the record may be released; … Identify specifically the intended uses of the information by the persons or organization; and … Have been published in the Federal Register. … If a Federal statute or an Executive Order of the President directs that records contained in a system of records be disclosed outside the Department of Defense, the statute or Executive Order serves as authority for the establishment of a routine use. … New or altered routine uses must be published in the Federal Register at least 30 days before any records may be disclosed pursuant to the terms of the routine use …." DoD 5400.11-R, *supra*, at ¶¶ C4.2.3.1, C4.2.3.2, C4.2.3.2.1, C4.2.3.2.2, C4.2.3.2.3, C4.2.3.2.4, C4.2.3.3, & C4.2.3.4; *see also id.* at ¶ DL1.21 (defining "routine use" as "[t]he disclosure of a record outside the Department of Defense for a use that is compatible with the purpose for which the information was collected and maintained by the Department of Defense. The routine use must be included in the published system notice for the system of records involved.")

[85] Nevertheless, for such purposes the Department of Justice holds that "[a] 'disclosure' can be by any means of communication—written, oral, electronic, or mechanical," and notes that guidelines from the Office of Management and the Budget (OMB) "and some, but not all, courts have advised that disclosures can occur by either transferring a record or simply 'granting access' to a record." "Overview of the Privacy Act," *supra*, from the "Conditions of Disclosure to Third Parties," *available at* https://www.justice.gov/opcl/overview-privacy-act-1974-2020-edition/disclosures-third-parties. Defense Department regulations, however, do provide an explicit definion of "disclosure." DoD 5400.11-R, *supra*, at ¶ DL1.5 ("The transfer of any personal information from a system of records by any means of communication (such as oral, written, electronic, mechanical, or actual review) to any person, private entity, or Government Agency, other than the subject of the record, the subject's designated agent, or the subject's legal guardian.").

[86] Agency-specific implementing guidelines generally track these basic statutory definitions. *See, e.g., See* DoD 5400.11-R, *supra*, at ¶¶ DL1.1.13 (definition of "record"), DL1.24 ("system of records"), DL1.8 ("individual"); ODNI, *Civil Liberties and Privacy Guidance, supra*, at 8 n.6 (conditions of disclosure) & 15 (definition of "system of records"); U.S. Department of Justice, "Private Act of 1974" (April 30, 2021) (definition of "system of records"), *available at* https://www.justice.gov/opcl/privacy-act-1974. (Note, however, that § G2 of DoDD 3115.18 ties its definition of "personally identifiable information" to that provided "in Section 552a of Title 5, United States Code, also known as the 'Privacy Act,' as amended," even though the phrase "personally identifiable information" does not actually appear in 5 U.S.C. § 552a.) Note also that it is generally not permitted to gather or retain information related to how an individual exercises any right protected by the First Amendment. 5 U.S.C. § 552a(e)(7); DoD 5400.11-R, *supra*, at ¶ C1.1.5.1; *see also* Henricks, *supra*, at 28.

[87] 5 U.S.C. § 552a(a)(5) (defining "system of records" as one "from which information is retrieved" in such fashion, rather than one from which information *can be* or *could be* thus retrieved). Defense Department regulations have drawn explicit attention to this distinction. *See* DoD 5400.11-R, *supra*, at ¶¶ C1.1.2.1, C1.1.2.1.1. & C1.1.2.1.2 ("Records in a group of records that may be retrieved by a name or personal identifier are not covered by this Regulation, even if the records contain personal data and are under control of a DoD Component. The records must be retrieved by name or other personal identifier to become a system of records for the purpose of this Regulation. … When records are contained in an automated Information Technology (IT) system that is capable of being manipulated to retrieve information about an individual, this does not automatically transform the system into a system of records, as defined in this Regulation. … In determining whether an automated system is a system of records that is subject to this Regulation, retrieval policies and practices shall be evaluated. If DoD Component policy is to retrieve personal information by name or other unique personal identifier, it is a system of records. If DoD Component policy prohibits retrieval by name or other identifier, but the actual practice of the Component is to retrieve information by name or identifier, even if done infrequently, it is a system of records."). *But see id.* at ¶ DL1.8 ("Corporations, partnerships, sole proprietorships, professional groups, businesses, whether incorporated or unincorporated, and other commercial entities are not 'individuals' when acting in an entrepreneurial capacity with the Department of Defense, but are 'individuals' when acting in a personal capacity (*e.g.,* security clearances, entitlement to DoD privileges or benefits, etc.).")

[88] It is also necessary that the requestor articulate "in sufficient detail why the records are required so that the custodian of the records may make an informed decision regarding their release," and that "[t]he intended use of the record generally relate[] to the purpose for which the record is maintained. *See* DoD 5400.11-R, *supra*, at ¶¶ C4.1.2, C4.2.1.1, C4.2.1.1.1, & C4.2.1.1.2. (Note also that for purposes of such intra-Departmental sharing, "[o]nly those records as are minimally required to accomplish the intended use are disclosed. The entire record is not released if only a part of the record will be responsive to the request." *Id.* at ¶ C4.2.1.1.3.)

[89] *Id.* at ¶ C1.3.1.1.

[90] *Id.* at ¶ C1.3.4.

[91] *See, e.g.,* DoD 5400.11-R, *supra*, at ¶¶ C4.2.3.1, C4.2.3.2, C4.2.3.2.1, C4.2.3.2.2, C4.2.3.2.3, C4.2.3.2.4, C4.2.3.3, C4.2.3.4, & DL1.21; Henricks, *supra*, at 28.

[92] Department of Defense Directive 5200.27 (January 7, 1980) [hereinafter "DoDD 5200.27"], at ¶¶ 2.2.1 & 2.2.2, *available at* https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/520027p.pdf.

[93] DoDD 5200.27, *supra*, at ¶ 3.1.

[94] *Id.* at ¶ 3.1.

[95] *Id.* at ¶ 4.1.

[96] *See Id.* at ¶¶ 4.1.1, 4.1.2, 4.1.4, 4.1.5, 4.1.7, 4.2, & 4.3.

[97] *Id.* at ¶¶ 4.1.3 & 4.1.6.

[98] U.S. Department of Defense, "Securing Defense-Critical Supply Chains: An action plan developed in response to President Biden's Executive Order 14017" (February 2022), at 3 & 10, *available at* https://media.defense.gov/2022/Feb/24/2002944158/-1/-1/1/dod-eo-14017-report-securing-defense-critical-supply-chains.pdf.

[99] Federal Acquisition Regulations (FAR), Part 10.001(a)(3) (noting that agencies "shall" use market research as specified), *available at* https://www.acquisition.gov/far/part-10.

[100] FAR, *supra*, at Part 10.002(b)(2)(iv).

[101] DoD, "Securing Defense-Critical Supply Chains," *supra*, at 26 & 54.

[102] E.O. 12333, *supra*, at § 2.3; *see also, e.g.,* David Kris, "The CIA's New Guidelines Governing Publicly Available Information," *Lawfare* (March 21, 2017) (summarizing Executive Order), *available at* https://www.lawfareblog.com/cias-new-guidelines-governing-publicly-available-information.

[103] E.O. 12333, *supra*, at § 2.7.

[104] *Id.* at § 3.5(k).

[105] The assiduous reader can find these various IC rule-sets online. *See* "Attorney General Approved U.S. Person Procedures Under E.O. 12333" (March 2021), *available at* https://www.intel.gov/assets/documents/guide/Chart_of_EO_12333_AG_approved_Guidelines_March_2021.pdf.

[106] E.O. 12333, *supra*, at § 2.3.

[107] Kris, *supra*.

[108] CIA Procedures, *supra*, at § 12.24.

[109] *See id.* at § 7 (allowing CIA to "retain information that has been lawfully collected concerning a U.S. person if … [t]he information is processed to delete USPII. In such cases, a generic term that does not identify the U.S. person in the context of the information, such as 'investor,' may be substituted.").

[110] *Id.* at § 8.2.

[111] *Id.* at § 12.25.

[112] *Id.*

[113] *Id.* at § 7. This section of the Procedures list the circumstances in which the CIA may retain information concerning a U.S. person that has been lawfully connected. The categories it lists, however, are provided in the disjunctive rather than the conjunctive—that is, they are connected by "or" rather than "and," thus signaling that each is an *alternative* exception to the general rule of not retaining U.S. person information. Subsection (a) specifies that one exception is where USPII has been removed, and subsection (b) specifies that another exception is where "[t]he information is publicly available."

[114] *Id.* at § 4.1 (noting that it is permissible to collect such information incidentally, if such collection is appropriately documented).

[115] *Id.* at § 2.3.

[116] *Id.* at § 4.2(a).

[117] *See id.* at at § 3.3 ("All duly authorized CIA activities subject to these Procedures shall have a purpose consistent with the CIA authorities described in Section 2. In any collection activity, the CIA shall collect only the amount of information reasonably necessary to support that purpose.").

[118] *Id.* at § 4.2.1. By comparison, collecting information concerning a U.S. person that isn't publicly available—but that may be obtained by means such as simply checking "existing records or knowledge of third parties (such as human sources, other federal agencies, or foreign governments)"—is termed "standard collection." *Id.* at § 4.3.1. (Standard collection may be undertaken only with the approval of specified senior officials. *See id.* at § 4.3.3.) "Special collection," in turn, is of a sort that "under the Fourth Amendment to the U.S. Constitution, would require a warrant if employed inside the United States for a law enforcement purpose." *Id.* at § 4.4. (Special collection requires even more elaborate approvals. *See id.* at § 4.4.2.)

[119] *Id.* at § 12.20.

[120] *Id.* at § 12.20. The same provision, however, also warns that "other commercial acquisitions of data may be so tailored and specialized for government use, and unavailable to a similarly situated private-sector purchaser, that the data cannot be considered publicly available."

[121] Kris, *supra* (internal citations omitted) (citing CIA Procedures, *supra*, at §§ 7(a)-(b) & 8.2.1(j)).

[122] CIA Procedures, *supra*, at § 6.2.3(a) & (c).

[123] *Id.* at § 8.2.2.

[124] *Id.* at § 12.2.

[125] *Id.* at § 5.1.

[126] *See, e.g., id.* at § 6.3.2(a) (noting that unevaluated information may be subjected to "routine handling requirements"—rather than "exceptional" ones—where that information "is stored in such a manner that it cannot be retrieved by reference to USPII").

[127] *See id.* at § 7(b).

[128] CIA Procedures, *supra*, at § 6.2.3.

[129] This may be less of a problem under the Privacy Act. As noted above, the Privacy Act deems a database to be a "system of records" covered by the Act where information is retrieved from it "by the name of [an] individual or by some identifying number, symbol, or other identifying particular assigned to the individual." 5 U.S.C. § 552a(a)(5). On its face, however, this would *see*m only to apply to *federal* government databases and not to cover queries made *by* federal officials (or their agents in the form of contractors) of databases independently maintained by private commercial data aggregators—though care would have to be taken to apply with Privacy Act rules with regard to how one stored and subsequently retrieved information about U.S. individuals that comes back from such queries of privately held sources.

[130] *See* Kris, *supra*.

[131] Department of Defense Manual 5240.01, *Procedures Governing the Conduct of DoD Intelligence Activities* (August 8, 2016) [hereinafter "DoDM 5240.01"], at § G2 (Glossary), *available at* https://dodsioo.defense.gov/Portals/46/DoDM%20%205240.01.pdf?ver=2016-08-11-184834-887. It does not, however, include "[a] reference to a product by brand or manufacturer's name or the use of a name in a descriptive sense, as, for example, Ford Mustang or Boeing 737," or "[i]magery from overhead reconnaissance or information about conveyances (e.g., vehicles, aircraft, or vessels) without linkage to additional identifying information that ties the information to a specific U.S. person." *Id.*

[132] DoDM 5240.01, *supra*, at § G2 (Glossary).

[133] *Id.* at ¶ 3.2.c. (This paragraph also lists other exceptions to the ban on intentionally collecting USPI, but these are not relevant for purposes of this paper.)

[134] *Id.* at § G2 (Glossary). The final reference to "information generally available to persons in a military community" seems to be unique to the DoD definition.

[135] *Id.* at ¶ 3.4.c.

[136] *Id.* at ¶ 3.2.f(2).

[137] *Id.* at ¶ 3.4.e.

[138] *Id.* at ¶ 3.4.d.

[139] *Id.* at ¶ 3.3.f(2).

[140] *Id.* at ¶ 3.2.c(1).

[141] *Id.* at ¶ 3.2.g(1) (noting that a Defense Intelligence Component "may only collect foreign intelligence concerning U.S. persons in the United States" if "the information is publicly available").

[142] *Id.* at ¶ 3.4.c(1).

[143] *Id.* at ¶ 3.4.c(3)-(5). There is even a provision for dissemination to a foreign government or international organization if "[t]he Defense Intelligence Component head or a delegee has determined that the disclosure is consistent with applicable international agreements and foreign disclosure policy and directives, including those policies and directives requiring protection against the misuse or unauthorized dissemination of information, and the analysis of potential harm to any individual." *Id.* at ¶ 3.4.c(6)(c).

[144] *Id.* at ¶¶ 3.3.e(1) (noting that "a Defense Intelligence Component may permanently retain USPI if it determines that retention is reasonably believed to be necessary for the performance of an authorized intelligence mission or function and the USPI falls into one or more of the following categories: (a) The information was lawfully collected by the Component or disseminated to the Component by another Component or element of the Intelligence Community and meets a collection category in Paragraph 3.2.c.") & 3.2.c (specifying that collection is permissible for information that is "publicly available").

[145] In fact, this is not merely a hypothetical case, as just such an effort is being developed at the MITRE Corporation and is being prepared for multi-stakeholder scalability.

[146] *See, e.g.,* 31 U.S.C. §§ 1301(a) [the "Purpose Statute"] (providing that "[a]ppropriations shall be applied only to the objects for which the appropriations were made except as otherwise provided by law") & 1341(1)(A) [the "Antideficiency Act"] (generally prohibiting government officials from "mak[ing] or authoriz[ing] an expenditure or obligation exceeding an amount available in an appropriation or fund for the expenditure or obligation"); U.S. Constitution, at Art. I, § 9, clause 7 ("No Money shall be drawn from the Treasury, but in Consequence of Appropriations made by Law.").

[147] This would not necessarily mean starting with the WON challenge that is most pressing. It might be more useful to prioritize the initial "target" in a way that balances its substantive importance with its *manageability* as a test case for data-use reforms. Biting off too sprawling a challenge at the outset could be counterproductive; it might be better to start with a slightly less important *but clearly achievable* goal and build toward applying such approaches to more-sweeping and "wickeder" problems as lessons are learned from the initial effort.

[148] To be sure, the provision might exempt certain agencies from these provisions. Health and Human Services, for instance, might be exempted to protect the absolute privacy of medical records, and likewise for the Internal Revenue Service (IRS) for tax filing information.

[149] *See* DoDD 3115.18, *supra,* at ¶¶ 2.1(d), 2.2(d), & 3.1.

[150] U.S. Department of Defense, "DOD Adopts Ethical Principles for Artificial Intelligence" (February 24, 2020), *available at* https://www.defense.gov/News/Releases/Release/Article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/. The five principles articulated for this purpose cover five major areas, as follows: (1) Responsible. DoD personnel will exercise appropriate levels of judgment and care, while remaining responsible for the development, deployment, and use of AI capabilities; (2) Equitable. The Department will take deliberate steps to minimize unintended bias in AI capabilities; (3) Traceable. The Department's AI capabilities will be developed and deployed such that relevant personnel possess an appropriate understanding of the technology, development processes, and operational methods applicable to AI capabilities, including with transparent and auditable methodologies, data sources, and design procedure and documentation; (4) Reliable. The Department's AI capabilities will have explicit, well-defined uses, and the safety, security, and effectiveness of such capabilities will be subject to testing and assurance within those defined uses across their entire life cycles; and (5) Governable. The Department will design and engineer AI capabilities to fulfill their intended functions, while possessing the ability to detect and avoid unintended consequences, and the ability to disengage or deactivate deployed systems that demonstrate unintended behavior. *Id.*

[151] Office of the Director of National Intelligence (ODNI), "Principles of Artificial Intelligence Ethics for the Intelligence Community" (undated), *available at* https://www.dni.gov/files/ODNI/documents/Principles_of_AI_Ethics_for_the_Intelligence_Community.pdf. This document declares that the IC "commits to the design, development, and use of AI with the following principles:" (1) Respect the Law and Act with Integrity. We will employ AI in a manner that respects human dignity, rights, and freedoms. Our use of AI will fully comply with applicable legal authorities and with policies and procedures that protect privacy, civil rights, and civil liberties; (2) Transparent and Accountable. We will provide appropriate transparency to the public and our customers regarding our AI methods, applications, and uses within the bounds of security, technology, and releasability by law and policy, and consistent with the Principles of Intelligence Transparency for the IC. We will develop and employ mechanisms to identify responsibilities and provide accountability for the use of AI and its outcomes; (3) Objective and Equitable. Consistent with our commitment to providing objective intelligence, we will take affirmative steps to identify and mitigate bias; (4) Human-Centered Development and Use. We will develop and use AI to augment our national security and enhance our trusted partnerships by tempering technological guidance with the application of human judgment, especially when an action has the potential to deprive individuals of constitutional rights or interfere with their free exercise of civil liberties; (5) Secure and Resilient. We will develop and employ best practices for maximizing reliability, security, and accuracy of AI design, development, and use. We will employ security best practices to build resilience and minimize potential for adversarial influence; and (6) Informed by Science and Technology. We will apply rigor in our development and use of AI by actively engaging both across the IC and with the broader scientific and technology communities to utilize advances in research and best practices from the public and private sector. *Id; see also* ODNI, "Artificial Intelligence Ethics Framework for the Intelligence Community" (June 2020), *available at* https://www.dni.gov/files/ODNI/documents/AI_Ethics_Framework_for_the_Intelligence_Community_10.pdf.

MITRE

SOLVING PROBLEMS
FOR A SAFER WORLD®