(U) Office of the Director of National Intelligence
Senior Advisory Group
Panel on Commercially Available Information

(U) Report to the Director of National Intelligence
27 January 2022

Classified By: [redacted]
Derived From: [redacted]
Declassify On: [redacted]

27 January 2022

The Honorable Avril Haines
Director of National Intelligence
Washington, DC 20511

Dear DNI Haines:

(U) With this letter, we transmit our 90-day report on commercially available information (CAI). We appreciate your commissioning the report and the assistance of your office and other Intelligence Community (IC) elements in this time-sensitive undertaking.

(U) As prescribed in our terms of reference (TOR), the report attempts to "(1) describe the role of CAI in intelligence collection and analysis; (2) reflect on the existing framework for ensuring the protection of privacy and civil liberties; and (3) make[] recommendations to the IC regarding how and under what circumstances an IC element should collect, use, retain, and disseminate CAI." These three issues, preceded by a background description and explanation of CAI, are addressed in the four main parts of our report.

(U) Our report does not attempt "an independent legal analysis" of the issues involved with CAI, as set forth in our TOR, but instead follows the IC's own approach in considering questions of CAI policy.

(U) Our report addresses CAI that is available for purchase by the general public and as such is treated as a subset of publicly available information (PAI). Unless otherwise indicated in context, we use the term "CAI" in this report to refer to CAI that is also PAI.

(U) Highlights of our report include the following:

1. (U) There is today a large and growing amount of CAI that is available to the general public, including foreign governments (and their intelligence services) and private-sector entities, as well as the IC.

2. (U) CAI clearly provides intelligence value, whether considered in isolation and/or in combination with other information, and whether reviewed by humans and/or by machines. It also raises significant issues related to privacy and civil liberties. The widespread availability of CAI regarding the activities of large numbers of individuals is a relatively new, rapidly growing, and increasingly significant part of the information environment in which the IC must function.

3. (U) Under IC elements' rules and procedures, CAI (because it is also PAI) is less strictly regulated than other forms of information acquired by the IC. In our view, however, profound changes in the scope and sensitivity of CAI have overtaken traditional understandings, at least as a matter of policy. Today's publicly available CAI is very different in degree and in kind from traditional PAI.

4. (U) We have three recommendations concerning the acquisition and treatment of CAI by the IC.

(U) First, the IC should develop a multi-layered process to catalog, to the extent feasible, the CAI that IC elements acquire. This will be a complex undertaking requiring attention to procurement contracts, functionally equivalent data acquisition processes, data flows, and data use. The IC cannot understand and improve how it deals with CAI unless and until it knows what it is doing with CAI.

(U) Second, based on the knowledge gained from that process, the IC should develop a set of standards and procedures for CAI, governing and requiring regular re-evaluation of acquisition and use decisions, including as to the use of CAI. We offer several points that can be included in those standards and procedures, but also recognize that they will need to be adapted for different IC elements with different CAI needs and missions.

(U) Third, as part of this set of policies and procedures, and/or as a complement to it, the IC should develop more precise guidance to identify and protect sensitive CAI that implicates privacy and civil liberties concerns. Again, we offer several suggestions for the development of such guidance.

(U) The single most important point in our report is this: CAI is increasingly powerful for intelligence and increasingly sensitive for individual privacy and civil liberties, and the IC therefore needs to develop more refined policies to govern its acquisition and treatment. Our report does not prescribe those policies (in keeping with our timeline and role as outside advisors) but we hope that it will assist the IC with their development.

(U) We appreciate the opportunity to be of service.

Respectfully submitted,

[redacted]

(U) **TABLE OF CONTENTS**

(U) EXECUTIVE SUMMARY

(U) There is today a large and growing amount of what the U.S. Intelligence Community (IC) refers to as "Commercially Available Information" (CAI). As the acronym indicates, and as we use the term in this report, CAI is information that is available commercially to the general public, and as such, is a subset of publicly available information (PAI). We do not use the term CAI to include, and we do not address in this report, commercial information that is available exclusively to governments. The volume and sensitivity of CAI have expanded in recent years mainly due to the advancement of digital technology, including location-tracking and other features of smartphones and other electronic devices, and the advertising-based monetization models that underlie many commercial offerings available on the Internet. Although CAI may be "anonymized," it is often possible (using other CAI) to deanonymize and identify individuals, including U.S. persons.

(U~~//FOUO~~) CAI clearly provides intelligence value, whether considered in isolation and/or in combination with other information, and whether reviewed by humans and/or by machines. The IC currently acquires a significant amount of CAI for mission-related purposes, including in some cases social media data [redacted] and many other types of information. As a resource available to the general public, including adversaries, CAI also raises counter-intelligence risks for the IC. It also has increasingly important risks and implications for U.S. person privacy and civil liberties, as CAI can reveal sensitive and intimate information about individuals. Without proper controls, CAI can be misused to cause substantial harm, embarrassment, and inconvenience to U.S. persons. The widespread availability of CAI regarding the activities of large numbers of individuals is a relatively new, rapidly growing, and increasingly significant part of the information environment in which the IC must function. That is the core of why it was necessary and appropriate for the IC to recognize the complex issues inherent in modern CAI and to commission this report.

(U~~//FOUO)~~ Under the U.S. Constitution, federal statutes, and IC elements' internal procedures, CAI is generally less strictly regulated than other forms of information acquired by the IC, principally because it is publicly available. In our view, however, changes in CAI have considerably undermined the historical policy rationale for treating PAI categorically as non-sensitive information, that the IC can use without significantly affecting the privacy and civil liberties of U.S. persons. For example, under *Carpenter v. United States*, acquisition of persistent location information (and perhaps other detailed information) concerning one person by law enforcement from communications providers is a Fourth Amendment "search" that generally requires probable cause. However, the same type of information on millions of Americans is openly for sale to the general public. As such, IC policies treat the information as PAI and IC elements can purchase it. While IC policies regulate such information based on the volume, proportion and sensitivity of USPI it contains, those policies may not accord sufficient

protection to information that is now broadly understood to be sensitive. It is not sufficient as a matter of policy simply to say that CAI is PAI; and saying so without more may be affirmatively confusing to intelligence professionals.

(U) We have three recommendations.

(U~~//FOUO~~) **First**, the IC should develop a multi-layered approach to **catalog, to the extent feasible, the acquisition and use of CAI across its 18 elements**. This cataloging process will be complex and should include formal contracts and procurement decisions, as well as functionally equivalent agency-specific data acquisition processes, because these will help identify CAI when it first arrives at (or becomes available to) an IC element. But the process also should include detection efforts at later stages of the information lifecycle, including in the process of planning for and initially using data. In particular, key inputs to the process may include (1) documentation reflecting the purchase, license, or other acquisition of a CAI dataset; (2) audits by chief information officers (CIOs) and chief data officers (CDOs) responsible for monitoring data flows across agency systems and repositories; and (3) [redacted] We recommend this multi-layered approach because prior retrospective efforts focused on procurement have not been successful, and because the dynamic nature of the CAI environment will require ongoing review. This first recommendation is foundational for our remaining two recommendations.

(U~~//FOUO~~) **Second**, as it gains knowledge into its own use of CAI, the IC should **develop a set of standards and procedures for CAI**, governing and requiring regular re-evaluation of acquisition and other decisions. This can be done centrally, for the IC or the Defense Intelligence Enterprise (DIE) as a whole, and/or at individual IC elements (where the approaches could vary from one element to another as long as they are consistent in principle). Either way, as the IC develops approaches to CAI, it will need to keep in mind IC elements' authorities and needs. Among the issues that should be considered in developing IC standards and procedures are the following: Mission analysis to identify need/value; Fit between mission and CAI data set, Proposed use; Vendor and data quality; Acquisition mechanics; Data security; Sensitivity and legal review; Auditing use of CAI; Periodic re-evaluation; and Other structural and procedural issues. We review several examples of IC elements' approaches to these issues, including the [redacted]

(U~~//FOUO~~) **Third**, as part of this set of standards and procedures, and/or as a complement to it, the IC should **develop more precise sensitivity and privacy-protecting guidance for CAI**. PAI is no longer a good proxy for non-sensitive information. Today, in a way that far fewer Americans seem to understand, and even fewer of them can avoid, CAI includes information on nearly everyone that is of a type and level of sensitivity that historically could have been obtained, if at

all, only through targeted (and predicated) collection, and that could be used to cause harm to an individual's reputation, emotional well-being, or physical safety. The IC therefore needs to develop more refined approaches to CAI. Among the structural and procedural issues that should be considered in developing such approaches are the following: Required involvement of relevant parties at all stages; VPS assessments generally being made prior to acquisition, or at least prior to analytic use of CAI, and ideally integrated or coordinated with CAI acquisition reviews discussed in Recommendation #2; Approval requirements, with higher levels of approval required for more sensitive cases; Documentation, retention, and availability to relevant personnel of assessments, approvals, and mitigation measures adopted; Re-evaluation of VPS assessments and measures; Forwarding of assessments and other documentation to ODNI and/or other central authorities, with a formal mechanism for periodic review to allow comparisons and discussion of best practices across IC elements and related purposes.

(U~~//FOUO)~~ Apart from the structural and procedural issues above, we recommend that IC elements also consider the following substantive issues in developing VPS guidance for CAI: Sensitivity of the CAI, in keeping with the discussion of sensitivity in Part 2 of this report; Deanonymization/reidentification issues; Importance of mission served by CAI (to balance against sensitivity of CAI); Strength of nexus between CAI and mission, and availability, feasibility, costs, and risks of alternatives; Ability to filter USPI prior to ingestion; Traditional minimization approaches and techniques; Availability of other privacy-protective measures in light of the need and anticipated use of CAI.

(U~~//FOUO)~~ Some IC elements have already made progress towards developing new VPS guidance, and we review several specific approaches that are in effect or are in the process of being developed. We offer four specific areas, drawn from the longer list above, in which such development would be particularly helpful. First, distinctions between types of CAI, including between historical CAI (e.g., newspapers) that are generally less sensitive, and newer forms of CAI that are generally more sensitive. Second, quantitative issues, because CAI that is acquired in bulk will almost always be more sensitive than CAI in smaller data sets. Third, special protections for USPs and USPI. Fourth and finally, issues raised by CAI that can easily be deanonymized, including implications for the definition of USPI as applied in this context.

(U) In **conclusion**, if some or all of our recommendations are agreeable, the IC will need a mechanism for putting them into effect – e.g., a traditional working group of IC senior officials. Such a working group might decide to proceed within the framework of our three recommendations, or it might adopt and build on their substance within a different framework – e.g., substantive principles; tools and procedures; and processes and approval requirements. We hope that our 90-day report provides a helpful foundation for developing more refined approaches, we believe that continued efforts will be necessary, and we appreciate the opportunity to be of service.

1. (U) BACKGROUND ON COMMERCIALLY AVAILABLE INFORMATION

(U) There is today a large and growing amount of what the U.S. Intelligence Community (IC) refers to as "Commercially Available Information" (CAI). As the acronym indicates, and as we use the term in this report, CAI is information that is available commercially to the general public, and as such, is a subset of publicly available information (PAI). We do not use the term CAI to include, and we do not address in this report, commercial information that is available exclusively to governments. The volume and sensitivity of CAI have expanded in recent years mainly due to the advancement of digital technology, including location-tracking and other features of smartphones and other electronic devices, and the advertising-based monetization models that underlie many commercial offerings available on the Internet. Although CAI may be "anonymized," it is often possible (using other CAI) to deanonymize and identify individuals, including U.S. persons.

1.1. (U) What is CAI? One of the challenges faced by the IC in dealing with CAI is defining the term, and hence the scope of any new guidance or policies that may be developed to address it. As the acronym indicates, and as we use it in this report, "CAI" is information that is available commercially, through a commercial transaction with another party. The acquisition may occur on a one-time or subscription basis, and may involve the IC directly ingesting the CAI or obtaining a license agreement that affords a continuing right of access. CAI typically is acquired for a fee, but as we use the term it also includes information offered at no cost if it is the type of information that is normally offered for sale – e.g., a free trial offering of CAI.

(U) As we use the term in this report, CAI does not include information that is stolen or otherwise misappropriated and then acquired from a black market or otherwise via traditional HUMINT acquisition methods (e.g., espionage). Nor does it include information obtained through traditional SIGINT acquisition methods (e.g., wiretapping) that does not involve a commercial transaction at all. As such, it does not necessarily include all information acquired from commercial entities, such as information acquired via lawful process (e.g., a search warrant or subpoena) served on a communications service provider or financial institution.

(U) In taking this approach to CAI, we generally follow the definition in the Intelligence Community Data Management Lexicon:

> (U) Any information that is of a type customarily made available or obtainable and sold, leased, or licensed to the general public or to non-governmental entities for purposes other than governmental purposes. Commercially Available Information also includes information for exclusive government use, knowingly and voluntarily provided by, procured from, or made accessible by corporate entities at the request of a government entity, or on their own initiative.

(U) Although some CAI is available only to governments, as the Lexicon notes, we use the term to mean, and this report addresses, only the subset of CAI that is generally available and is therefore also publicly available information (PAI). Under IC guidelines, 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 any casual observer (but not amounting to physical surveillance), is made available at a meeting open to the public, or is observed by visiting any place or attending any event that is open to the public.

(U) Office of the Director of National Intelligence, Intelligence Activities Procedures Approved by the Attorney General Pursuant to Executive Order 12333 (ODNI Guidelines) § 10.17 (emphasis added); see also Central Intelligence Agency Activities: Procedures Approved by the Attorney General Pursuant to Executive Order 12333 (CIA Guidelines) § 12.20; DOD Manual 5240.01: Procedures Governing the Conduct of DOD Intelligence Activities (DOD Manual) § G.2 at page 53.

(U) To repeat for emphasis and clarity, unless otherwise indicated in context, we use the term "CAI" to refer to CAI that is also PAI, and our report addresses only CAI that is also PAI. Non-public CAI raises distinct legal and policy questions and is beyond the scope of our current efforts.

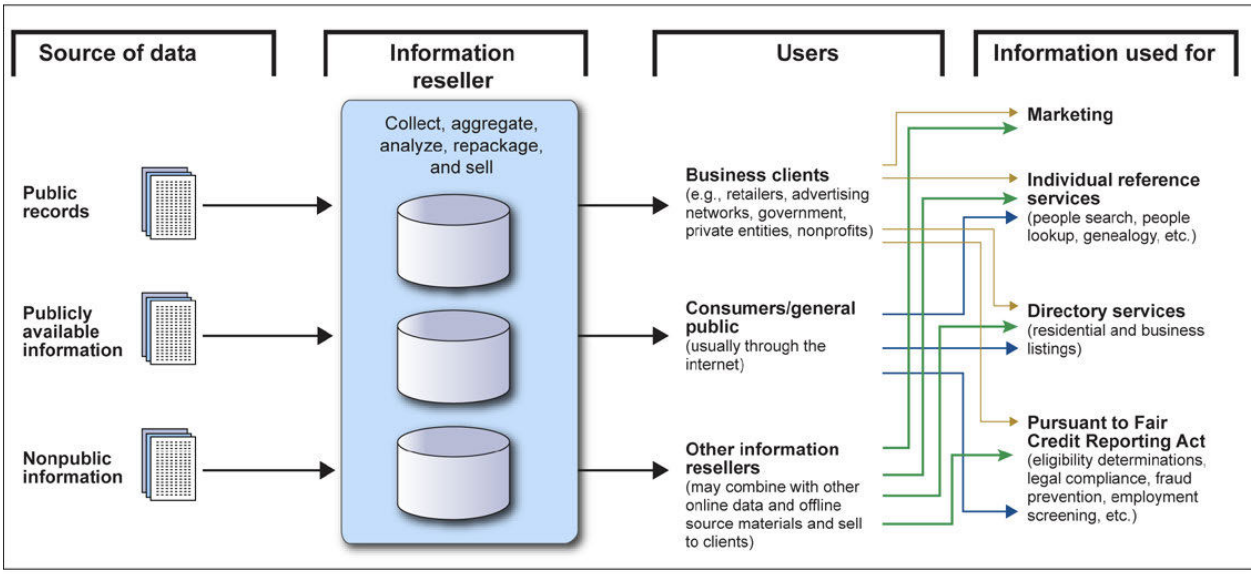
(U~~//FOUO~~) Our discussions with IC elements included a heavy emphasis on defining CAI, a valuable and foundational effort for this report and any future regulation of CAI. Focus on a formal definition, however, should not obscure the functional perspective that animates our recommendations. As discussed in the balance of this report, CAI merits special attention today because of its increasing importance for intelligence as well as for privacy and civil liberties (as discussed in Part 2), and because it has, at least in part, overtaken current IC policies that address it (as discussed in Part 3). Those concerns should inform efforts to apply, and if necessary to modify, the formal definition of CAI in the many, varied and evolving contexts that the IC does and will face. Cf. Privacy and Civil Liberties Oversight Board, Report to the President on Implementation of Presidential Policy Directive 28: Signals Intelligence Activities at 12, 24 (noting the absence of a formal definition of "signals intelligence" under PPD-28).

1.2. (U) CAI Sellers. A key feature of CAI is that it is often sold or otherwise made available by commercial entities. Sellers of CAI are often referred to as "data brokers" or "information resellers." As the Government Accountability Office (GAO) reported in December 2013, these sellers of CAI

maintain large, sophisticated databases with consumer information that can include credit histories, insurance claims, criminal records, employment histories, incomes, ethnicities, purchase histories, and interests. Resellers largely obtain their information from public records, publicly available information (such as directories and newspapers), and nonpublic information (such as from retail loyalty cards, warranty registrations, contests, and web browsing). Characterizing the precise size and nature of the reseller industry can be difficult because of limited publicly known information about the industry.

(U) In testimony before the Senate Banking Committee in June 2019, a GAO official repeated the substance of those observations from 2013 and provided the following graphic to illustrate the development of and market for CAI:

Graph is Unclassified



Source: GAO. | GAO-19-621T

(U) A May 2014 report from the Federal Trade Commission (FTC) provides a similar account:

(U) Data brokers collect data from commercial, government, and other publicly available sources. Data collected could include bankruptcy information, voting registration, consumer purchase data, web browsing activities, warranty registrations, and other details of consumers' everyday interactions. Data brokers do not obtain this data directly from consumers, and consumers are thus largely unaware that data brokers are collecting and using this information. While each data broker source may provide only a few data elements about a consumer's activities, data brokers can put all of these data elements together to form a more detailed composite of the consumer's life.

(U) Civil society groups in the United States have also described data brokers and the market for CAI in the context of their public advocacy efforts. A recent example is the report from the

Center for Democracy and Technology (CDT), *Legal Loopholes and Data for Dollars*, released in December 2021. As of this writing, major data brokers include Accenture, Acxiom, CoreLogic, Epsilon, Intelius, LexisNexis, Oracle (Datalogix), Thomson Reuters, and Verisk (these companies, and the ones further discussed below, are listed solely for purposes of illustration, and references to the work of civil society groups are similarly for descriptive purposes only). In general, CAI sellers include those focused on marketing and advertising, fraud detection, risk mitigation, and identity resolution (people finders). Purchasers of CAI include other data brokers, various private-sector and non-governmental entities, and governments worldwide, including the IC.

1.3. (U) Examples of CAI. We do not attempt a comprehensive description of the scope and scale of data that are available as CAI, or the relevant markets, in part because they are so large and so dynamic. However, a few examples of CAI offerings will illustrate the current nature of available offerings:

- (U) "Thomson Reuters CLEAR® is powered by billions of data points and leverages cutting-edge public records technology to bring all key content together in a customizable dashboard."

- (U) LexisNexis offers more than "84B records from 10,000+ sources, including alternative data that helps surface more of the 63M unbanked/underbanked U.S. adults."

- (U) Exactis has "over 3.5 billion records (updated monthly)" in its "universal data warehouse."

- (U) PeekYou "collects and combines scattered content from social sites, news sources, homepages, and blog platforms to present comprehensive online identities."

(U) As these examples show, there is a large and growing amount of CAI in existence and offered for sale, some of it sensitive with respect to privacy. The market for CAI, including analysis and exploitation of CAI for insight, is evolving both qualitatively (e.g., as to types of data available) and quantitatively (as to amounts of data available) – see, for example, this March 2021 summary from Gartner. It includes significant information on U.S. persons, much of which can be acquired in bulk. As discussed below, moreover, certain CAI that is "anonymized" and available in bulk can readily be reidentified to reveal information about individuals.

1.4. (U) Origins & Evolution of CAI. In substantial part, the vast and growing amount of available CAI results from evolving digital technology, and the proliferation of digital dust created by individuals in their daily lives. As our TOR explain, "[t]he digital revolution has placed an incredible amount of information into the hands of private actors, many of whom seek to sell the data." For example, CAI can be obtained from public records, sometimes digitized from

paper originals, such as information about real estate transactions that can be found in local title offices or courthouses. It can be obtained from smartphone and other software applications, often in the form of software development kits (SDK), that collect information from devices in the U.S. and abroad. And CAI can be obtained from cookies and other methods, sometimes associated with real-time bidding (RTB) for sales of online advertising, that track end users as they browse the Internet. In April 2021, a bipartisan group of U.S. Senators raised questions about "the sharing of Americans' data through 'real time bidding' – the auction process used to place many targeted digital advertisements." The details of these digital developments are beyond the scope of this report; it is sufficient for our purposes, and widely understood among intelligence professionals and policymakers, that they have significantly contributed to the profound increase in CAI.

1.5. (U) Commercial Value of CAI. Various forms of CAI can be combined to synergistic effect in service of various commercial interests. For example, according to an October 2020 press release from Gartner, the "internet of behaviors (IoB) is emerging as many technologies capture and use the 'digital dust' of peoples' daily lives. The IoB combines existing technologies that focus on the individual directly – facial recognition, location tracking and big data for example – and connects the resulting data to associated behavioral events, such as cash purchases or device usage." As the FTC explained in its May 2014 report:

> (U) Data brokers rely on websites with registration features and cookies to find consumers online and target Internet advertisements to them based on their offline activities. Once a data broker locates a consumer online and places a cookie on the consumer's browser, the data broker's client can advertise to that consumer across the Internet for as long as the cookie stays on the consumer's browser. Consumers may not be aware that data brokers are providing companies with products to allow them to advertise to consumers online based on their offline activities. Some data brokers are using similar technology to serve targeted advertisements to consumers on mobile devices.

(U) The commercial value of CAI is plainly high, which is why the market for CAI is large and growing.

1.6 (U) Deanonymization/Reidentification. CAI can also be combined, or used with other non-CAI data, to reverse engineer identities or deanonymize various forms of information. As the *New York Times* reported in December 2019, "[i]f you own a mobile phone, its every move is logged and tracked by dozens of companies … The Times Privacy Project obtained a dataset with more than 50 billion location pings from the phones of more than 12 million people in this country. It was a random sample from 2016 and 2017, but it took only minutes — with assistance from publicly available information — for us to deanonymize location data." The *Times* was able to track the movements of President Trump via a member of his Secret Service detail. Deanonymized data may be useful for commercial and/or intelligence purposes.

2. (U) THE ROLE OF CAI IN INTELLIGENCE COLLECTION AND ANALYSIS

(U//~~FOUO~~) *CAI clearly provides intelligence value, whether considered in isolation and/or in combination with other information, and whether reviewed by humans and/or by machines. The IC currently acquires a significant amount of CAI for mission-related purposes, including in some cases social media data* [redacted] *and many other types of information. As a resource available to the general public, including adversaries, CAI also raises counter-intelligence risks for the IC. It also has increasingly important risks and implications for U.S. person privacy and civil liberties, as CAI can reveal sensitive and intimate information about individuals. Without proper controls, CAI can be misused to cause substantial harm, embarrassment, and inconvenience to U.S. persons. The widespread availability of CAI regarding the activities of large numbers of individuals is a relatively new, rapidly growing, and increasingly significant part of the information environment in which the IC must function. That is the core of why it was necessary and appropriate for the IC to recognize the complex issues inherent in modern CAI and to commission this report.*

2.1. (U) CAI as a Source for OSINT. For the IC, CAI provides intelligence value as a form of publicly available information used to create Open Source Intelligence (OSINT), as well as for other purposes including force protection and enrichment of information in other INT disciplines. Many recent commissions and reports have focused on the value of CAI and other PAI as OSINT. For example, in 2005, the WMD Commission's report concluded (pages 22-23) that "analysts who use open source information can be more effective than those who don't," and urged creation of an "entity that collects, processes, and makes available to analysts the mass of open source information that is available in the world today." Similarly, the January 2019 "AIM Initiative" from the Office of the Director of National Intelligence (ODNI), which is a strategy for augmenting intelligence using machines, explained "the IC must develop both the capability and capacity to take advantage of available data across all INTs and open source, and develop AI solutions that process and relate information from multiple modalities." A January 2021 report from the Center for Strategic and International Studies (CSIS) notes that the IC must "encourage IC agencies to integrate OSINT into collection and analytic tradecraft," because the "combination of cloud, cloud-based AI and analytics tools, and commercial GEOINT and SIGINT collection means that high-quality, multi-source intelligence analysis can be produced at the unclassified level from anywhere equipped to do so."

2.2. (U) Examples of CAI Contracts. The IC currently acquires a large amount of CAI. Unclassified IC and other contracts for CAI can be found at Sam.Gov, a U.S. government website that allows searching by agency or sub-agency and by keywords, among other things. By way of example only, this website shows that the following agencies have, have had, have considered, or are considering the following contracts or proposals related to CAI:

- (U) The Federal Bureau of Investigation (FBI) with ZeroFox for social media alerting (15F06721P0002431)

- [redacted]

- (U) The Defense Intelligence Agency (DIA) for social media reports on individuals who are seeking a security clearance (HHM402-16-SM-CHECKS), and with LexisNexis for "retrieval of comprehensive on-line search results related to commercial due diligence from a maximum number of sources (news, company, public records, legal, regulatory financial, and industry information)," among other things (HHM402-21-Q-0094)

- (U) The U.S. Navy with Sayari Analytics, Inc. for access to its database that "contains tens of thousands of previously-unidentified specific nodes, facilities and key people related to US sanctioned actors including '2+3' threats to national security" (N0001518PR11212)

- (U) Various offices within the Treasury Department for access to Banker's Almanac (RFQ-FIN-55100-21-0010)

- (U) The Department of Defense (DOD) for access to Jane's online (W31P4Q17T0009)

- (U) The Coast Guard with Babel Street for "Open Source Data Collection, Translation, Analysis Application" (70Z08419QVA044).

(U) In addition, DIA has provided the following information about a CAI contract in an unclassified and publicly-available paper sent to Congress on January 15, 2021:

> (U) DIA currently provides funding to another agency that purchases commercially available geolocation metadata aggregated from smartphones. The data DIA receives is global in scope and is not identified as "U.S. location data" or "foreign location data" by the vendor at the time it is provisioned to DIA. DIA processes the location data as it arrives to identify U.S. location data points that it segregates in a separate database. DIA personnel can only query the U.S. location database when authorized through a specific process requiring approval from the Office of General Counsel (OGC), Office of Oversight and Compliance (OOC), and DIA senior leadership. Permission to query the U.S. device location data has been granted five times in the past two-and-a-half years for authorized purposes.

(U) In the process of preparing this report, DHS described for us three ways in which CAI is generally used by its Office of Intelligence and Analysis (I&A):

- (U~~//FOUO~~) Web of Science is a powerful targeting tool, as it allows DHS I&A analysts to quickly and efficiently search and triage a large repository of academic publications and filter according to funding sources, affiliations, co-authors, and other key terms. This

service provides critical insight into academic publications that are not easily found elsewhere or are hidden behind paywalls when searched via other means, such as Google Scholar. For example, using Web of Science, DHS I&A analysts have identified foreign researchers studying in the United States with previously unknown associations with their home country's military. Additionally, through the currently OSDLS-managed subscriptions, we are also able to access the Web of Science API, which allows us to apply data analytics to the database.

- (U~~//FOUO~~) CLEAR enables DHS I&A to resolve identities and also provides leads for further analysis in DHS systems, thereby focusing resources on threat actors and not innocent persons. Commercial databases like CLEAR often have current location and contact information as well. Often, given the target set we focus on – non-traditional collectors – intelligence collection is not sufficient to resolve identities of subjects of interest. Data available in a commercially available datasets enables identity resolution by comparing it to what's in DHS systems and also reduces the risk of misidentification.

- (U~~//FOUO~~) Dun and Bradstreet and similar tools enable DHS I&A to resolve private companies' primary enterprises with their subsidiaries/affiliates and provides leads for further analysis in DHS systems and classified databases. Access to this information is critical to countering malign foreign investment that may threaten the security and resiliency of U.S. critical infrastructure.

2.3. (U) Examples of CAI Value. In our classified briefings with IC elements, we discussed the intelligence value of CAI, including how it can be used to reduce cost and risk of acquisition that might otherwise occur through clandestine means. The IC is strongly of the view that it will be at a significant disadvantage vis a vis foreign adversaries and competitors if it does not enjoy certain access to CAI. We urge the IC to make available several unclassified examples showing the value of CAI because we believe it will help inform the policy debate, in keeping with Principles of Intelligence Transparency for the IC. The IC has done this in the past in other contexts, including for Section 702 of the FISA Amendments Act.

(U) Here are two unclassified examples provided by the IC in response to our request while we were preparing this report:

- (U~~//FOUO~~) "NSA's Cybersecurity Collaboration Center leveraged commercial and SIGINT sources to expand the community sight picture on the advanced persistent threat Cobalt Strike actor. Analysts used enterprise access from [redacted] and others to identify a pattern in the registration of the seed nodes shared by the 370 domains – of which 19 were tagged by [redacted] and of those, 7 resolved to CobaltStrike infrastructure. A pattern in uniform resource locators (URLs) was also discovered to be associated with CobaltStrike using CAI which led to the discovery of an additional 49 internet protocol (IP) addresses."

- (U~~//FOUO~~) "CAI allows the IC to create valuable products for excluded missions like HADR [humanitarian assistance and disaster response]. These products are similar to those created by the commercial and academic community. These types of use cases focus on providing strategic level analytic outputs on how events affect human mobility at-scale or the country level. Examples include but are not limited to, how natural disasters and the spread of disease affect the movement of humans and vice versa."

(U) We expect that the IC will be able to provide additional unclassified examples over time. If necessary, moreover, classified examples should be made available to appropriate audiences. We believe that CAI is extremely and increasingly valuable and important for the conduct of modern intelligence activity, both as a source of OSINT and to support, enrich and enable other INT disciplines.

2.4. (U) Non-Analytic Uses of CAI. It is important to recognize that in some cases, CAI may also be used for purposes other than intelligence collection and analysis. At the outset, of course, the FBI uses CAI under its law enforcement authorities, as authorized in AG Guidelines and FBI policy, for the investigation of criminal matters, and non-intelligence elements of DOD may also use CAI for their missions. We also briefly consider three non-analytic intelligence use cases of CAI.

(U~~//FOUO~~) First, CAI may be useful in supporting compliance with legal or policy requirements. For example, geolocation CAI might be able to support compliance with 50 U.S.C. § 1881a (Section 702 of the FISA Amendments Act), which generally applies only to collection targeting non-U.S. persons reasonably believed to be located outside the United States. CAI can help determine location for compliance with this core requirement of Section 1881a. It may also be useful in complying with requirements established by Congress for situations in which non-U.S. persons abroad who are under surveillance travel into the United States. See 50 U.S.C. § 1805(f). More generally, CAI may also help establish the "foreignness" of SIGINT or other collection targets as necessary to meet legal or policy requirements.

(U~~//FOUO~~) Second, CAI may also be used in support of clandestine and HUMINT operations. CAI utilized in support of operations uniquely enables activities like cover development and operations planning. These activities are tightly held within the IC and subject to extremely restrictive operations access and handling rules. Further, CAI data obtained to support operations is outside the IC's classic analysis and intelligence reporting streams – it is not disseminated.

(U) Third, CAI may be useful in building and training artificial intelligence models. Although non-analytical in the strict sense, such models themselves can then be used to gain analytic insight or for other purposes.

(U) We do not mean to suggest a policy outcome in these particular use cases. Our only point is that policy questions concerning CAI are not one-dimensional. The importance and nature of the need, the absence of viable alternatives for meeting it, restrictions on access to and use and dissemination of data, should all be considered in reaching an appropriate policy judgment in each case – assuming, as always, that outcomes are not dictated by law. It may be, for example, that certain privacy-protecting methods, such as encrypting, masking, and use of differential privacy, may be viable for some mission needs even if not for others.

2.5. (U) Counter-Intelligence Risks in CAI. There is also a growing recognition that CAI, as a generally available resource, offers intelligence benefits to our adversaries, some of which may create counter-intelligence risk for the IC. For example, the January 2021 CSIS report cited above also urges the IC to "test and demonstrate the utility of OSINT and AI in analysis on critical threats, such as the adversary use of AI-enabled capabilities in disinformation and influence operations." Additional risks are developed in this April 2021 Lawfare article and this August 2021 report from Duke University.

(U) The Duke University report describes certain counter-intelligence risks from CAI. It finds, for example, that of 10 major data brokers surveyed, three advertise that they can provide data to identify U.S. military personnel. The report goes on to note (as summarized in a Lawfare article by its author): "Foreign actors could use this data to bolster their influence campaigns to interfere in U.S. electoral processes. Criminal organizations could use this data to build profiles on and subsequently target prosecutors and judges. Foreign intelligence organizations could acquire this data through a variety of means—including through front companies that could legally purchase the data from U.S. brokers and through simply hacking a data broker and stealing it all—to build profiles on politicians, media figures, diplomats, civil servants, and even suspected or secretly identified intelligence operatives."

(U) We have not necessarily validated these examples with the IC, meaning that they should not necessarily be taken as unresolved risks; but they illustrate the types of risk that CAI can create in the hands of our adversaries.

2.6. (U) Sensitivity of CAI. CAI can reveal sensitive and intimate information about the personal attributes, private behavior, social connections, and speech of U.S. persons and non-U.S. persons. It can be misused to pry into private lives, ruin reputations, and cause emotional distress and threaten the safety of individuals. Even subject to appropriate controls, CAI can increase the power of the government's ability to peer into private lives to levels that may exceed our constitutional traditions or other social expectations. Mission creep can subject CAI collected for one purpose to other purposes that might raise risks beyond those originally calculated. The IC's use of CAI is also the subject of intense scrutiny and speculation by political leaders, the news media, and civil society.

2.6.1. (U) CAI Includes Sensitive and Intimate Information. CAI can contain information that is deemed sensitive, meaning information that is not widely known about an individual that could be used to cause harm to the person's reputation, emotional well-being, or physical safety. As a primary justification for finding precise cell-site location information subject to Fourth Amendment protection in *Carpenter v. United States*, 138 S. Ct. 2206 (2018), the Supreme Court focused on how the "data provides an intimate window into a person's life, revealing not only his particular movements, but through them his 'familial, political, professional, religious, and sexual associations.' These location records 'hold for many Americans the 'privacies of life.'" *Id*. at 2217. CAI can also contain intimate information, meaning information that reveals private details about how people relate to one another.

2.6.2. (U) Defining Sensitivity Categorically. Many statutes, rules, and privacy policies describe information sensitivity categorically, listing types of information that tend to raise risks of harm. To give a comparative example, the European General Data Protection Regulation (GDPR) identifies as sensitive:

> (U) personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership., and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation.

(U) Similarly, an internal data call from ODNI listed several categories of CAI as potentially sensitive, including:

> (U) persistent location information, medical (to include mental health) information, travel records, attorney-client information, information concerning … religion or religious practices, information containing data on sexual activity, records regarding purchases, library records [as well as information] regarding individuals' communications [metadata and content] [and] information concerning individuals' expression of ideas or political views or the groups or individuals with whom they associate.

2.6.3. (U~~//FOUO~~) CAI Can Be Misused. Studies document the extent to which large collections of sensitive and intimate information about individuals, CAI or not, can be subject to abuse. Documented examples of LOVEINT abuses (government officials spying on actual or potential romantic partners) involving other intelligence collections demonstrate the potential for comparable abuse of CAI held by the IC. In the wrong hands, sensitive insights gained through CAI could facilitate blackmail, stalking, harassment, and public shaming. Concerns like these are why, as detailed in Part 4, several IC elements require a "volume, proportion, and sensitivity" analysis of certain data practices that considers, among other things, the "potential for substantial harm, embarrassment, inconvenience, or unfairness to U.S. persons if the information is improperly used or disclosed."

2.6.4. (U) CAI Increases the Power of the Government. The government would never have been permitted to compel billions of people to carry location tracking devices on their persons at all times, to log and track most of their social interactions, or to keep flawless records of all their reading habits. Yet smartphones, connected cars, web tracking technologies, the Internet of Things, and other innovations have had this effect without government participation. While the IC cannot willingly blind itself to this information, it must appreciate how unfettered access to CAI increases its power in ways that may exceed our constitutional traditions or other societal expectations.

(U) CAI also implicates civil liberties. CAI can disclose, for example, the detailed movements and associations of individuals and groups, revealing political, religious, travel, and speech activities. CAI could be used, for example, to identify every person who attended a protest or rally based on their smartphone location or ad-tracking records. Civil liberties concerns such as these are examples of how large quantities of nominally "public" information can result in sensitive aggregations.

2.6.5. (U) Aggregation of CAI Raise the Risk of Mission Creep. CAI collected for one purpose may be reused for other purposes. An assessment of the risk to privacy of data collected at one point in time may differ materially from a reassessment of the risk as applied to new purposes.

2.6.6. (U) Public, Media, and Political Scrutiny. The public seems to care about the risk to personal privacy posed by the accumulation and sale of personal information by online platforms, smartphone apps, connected devices, and other commercial entities. A steady string of public controversies including Cambridge Analytica, the revelations by the *New York Times* about data brokers that sell location information, the use of app usage data to identify a priest who was using the Grindr app, and the revelation of the sale of usage data by a Muslim prayer app, among many other examples, demonstrate the keen interest in CAI, at least on the part of the media, civil society groups, and political leaders. The possible future revelation that any component of the IC has gathered CAI without a proper accounting for the costs and benefits raises the risk of significant media attention and political fallout and could jeopardize other forms of CAI collection and use.

2.6.7. (U) Need for Thoughtful Approach. None of this is to suggest that CAI should be categorically off-limits to the IC; the CAI that we address is publicly available, including to friendly and adversarial foreign governments (and their intelligence services), non-governmental organizations, commercial entities of many kinds, and individuals. It is only to say that the privacy and civil liberties concerns that underlie judicial decisions like *Carpenter*, and possible legislation restricting access to CAI, are real and important, and that the IC should therefore take responsibility to develop a thoughtful and balanced approach in this area.

(U) As noted above, we think it is insufficient as a matter of policy to treat all CAI as PAI, without more, because modern CAI is so different from traditional PAI. Today's CAI is more revealing, available on more people (in bulk), less possible to avoid, and less well understood than traditional PAI. It is only a little oversimplified to say that when Executive Order 12333 was adopted, U.S. persons generally understood that the White Pages and the *New York Times* were public, but also understood that it was possible to choose an unpublished telephone number and (usually) to keep oneself out of the newspaper. Today, in a way that far fewer Americans seem to understand, and even fewer of them can avoid, CAI includes information on nearly everyone that is of a type and level of sensitivity that historically could have been obtained, if at all, only through targeted (and predicated) collection. As a matter of policy, therefore, asserting that modern CAI is materially indistinguishable from traditional PAI "is like saying a ride on horseback is materially indistinguishable from a flight to the moon." *Riley v. California*, 573 U.S. 373, 393 (2014). These new qualitative and quantitative aspects of CAI, particularly of or concerning U.S. persons and as discussed in Section 4.3.4, are key sensitivity concerns that animate the need for a new approach.

2.7. (U) Summary. We have no doubt that CAI can provide significant intelligence value, both to the IC and to our adversaries, whether standing alone or in combination with other information that is collected using classified sources and methods, and whether analyzed by humans and/or by machines. It also clearly raises significant issues of privacy and sensitivity, including for U.S. persons. CAI is a relatively new, rapidly growing, and increasingly significant part of the information environment in which the IC must function. That is the core of why this report is necessary.

3. (U) THE EXISTING POLICY FRAMEWORK FOR CAI

(U) *Under the U.S. Constitution, federal statutes, and IC elements' internal procedures, CAI is generally less strictly regulated than other forms of information acquired by the IC, principally because it is publicly available. In our view, however, changes in CAI have considerably undermined the historical policy rationale for treating PAI categorically as non-sensitive information, that the IC can use without significantly affecting the privacy and civil liberties of U.S. persons. For example, under* Carpenter v. United States*, acquisition of persistent location information (and perhaps other detailed information) concerning one person by law enforcement from communications providers is a Fourth Amendment "search" that generally requires probable cause. However, the same type of information on millions of Americans is openly for sale to the general public. As such, IC policies treat the information as PAI and IC elements can purchase it. While IC policies regulate such information based on the volume, proportion and sensitivity of USPI it contains, those policies may not accord sufficient protection to information that is now broadly understood to be sensitive. It is not sufficient as a matter of policy simply to say that CAI is PAI; and saying so without more may be affirmatively confusing to intelligence professionals.*

3.1. (U) PAI. Historically, PAI has not been considered sensitive, as reflected in both U.S. law and policy. In keeping with our TOR, we do not offer an independent legal analysis of this issue; instead, we review the legal background governing PAI solely as context for our policy discussion of CAI.

3.1.1. (U) Constitutional Provisions. As a general matter, under the Fourth Amendment, "[w]hat a person knowingly exposes to the public, even in his own home or office, is not a subject of Fourth Amendment protection." *Katz v. United States*, 389 U.S. 347, 351 (1967). To be sure, more recent decisions, most notably *Carpenter v. United States*, 138 S. Ct. 2206 (2018), raise questions about the extent to which providing information to certain third parties can extinguish a reasonable expectation of privacy in that information. In keeping with our TOR, we do not attempt to answer those questions; it is enough for our purposes to recognize the general rule that PAI is not deemed sensitive. To take an obvious example, the Justices' signed opinions in *Carpenter* are clearly not protected by the Fourth Amendment and are available in searchable CAI data sets from Lexis, Westlaw, and other providers. Historically, PAI also generally was not considered sensitive under the First Amendment; but PAI today, including CAI, may implicate First Amendment rights.

3.1.2. (U) Federal Statutes. Resting on the constitutional understanding discussed above, many federal statutes expressly decline to protect, and assume the absence of Fourth Amendment protection for, PAI. The federal Wiretap Act, 18 U.S.C. § 2511(2)(g)(i), provides that it "shall not be unlawful . . . for any person . . . to intercept or access an electronic communication made through an electronic communication system that is configured so that such electronic communication is readily accessible to the general public." Similarly, the Foreign Intelligence

Surveillance Act defines "electronic surveillance" in ways that expressly incorporate Fourth Amendment principles and law enforcement standards, 50 U.S.C. §§ 1801(f), and requires "minimization" of "nonpublicly available information," 50 U.S.C. § 1801(h). To be sure, the Privacy Act, 5 U.S.C. § 552a, also places certain restrictions on the IC when it collects and retrieves USPI, including when the USPI is also PAI. The IC must have clear authority and mission need to collect this PAI; must provide a notice to the public about the collection (system of records notices), and generally may not maintain a record describing how an individual exercises First Amendment rights. Some PAI can include protected speech (e.g., social media posts) or associational information, and the Privacy Act would need to be considered before collecting such information.

3.1.3. (U) Pending Legislation. We are aware that there are federal legislative efforts underway that might affect the treatment of CAI, at least as acquired by the IC or other governmental entities. We do not express an opinion on the merits of any particular pending or contemplated legislation, but as an institutional matter we believe that legislation could address policy concerns with the current regulatory framework governing CAI.

3.1.4. (U) IC Policy. The foundational document governing the IC also treats PAI as relatively unprotected. Section 2.3 of Executive Order 12333 authorizes IC elements "to collect, retain, or dissemination information about U.S. persons" in accordance with procedures established by the head of the IC element and the Attorney General, and provides that these procedures "shall permit collection, retention, and dissemination of . . . [i]nformation that is publicly available or collected with the consent of the person concerned." The term "publicly available" is defined in the procedures of several IC elements. For example, the procedures for the Central Intelligence Agency (CIA) define "publicly available information" is as follows:

> (U) [1] information that has been published or broadcast for public consumption, [2] is available on request to the public, [3] is accessible online or otherwise to the public, [4] **is available to the public by subscription or purchase**, [5] could be seen or heard by any casual observer (but not amounting to physical surveillance), [6] is made available at a meeting open to the public, or [7] is obtained by visiting any place or attending any event that is open to the public.
>
> (U) Information is publicly available only if it is made available to the CIA under conditions or on terms generally applicable to the public. For example, certain commercially acquired data may be considered publicly available if a non-U.S. government person or corporation could acquire that same data in that same way from that same commercial source; however, 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.

(U) CIA Guidelines § 12.20 (emphasis added). The corresponding guidance for DOD intelligence elements adopts a similar definition and adds the clarification that "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." DOD Manual 5240.01 §§ 3.2.b and G-2 at page 53.

(U) Under Executive Order 12333, moreover, some IC elements are authorized to collect information (mainly or exclusively) "overtly or through publicly available sources." EO 12333 §§ 1.7(i)-(j), 1.8, 1.9, 1.12(c). The expansion of PAI to include modern CAI is highly consequential for the work of these IC elements.

3.1.5. (U) IC Guidance. In July 2011, ODNI issued Civil Liberties and Privacy Guidance for Intelligence Community Professionals: Properly Obtaining and Using Publicly Available Information. This guidance included a "shorthand, non-exhaustive list of factors to consider for properly obtaining and using" PAI, including that the PAI is (1) available to the general public; (2) lawfully obtained by the IC (e.g., if a hacker posts instructions on a blog for how to penetrate a bank's online security, the bank's data does not become lawfully available as a result); (3) the IC purchaser has complied with any requirements to disclose IC affiliation, which is often addressed through guidelines on "undisclosed participation" and similar rules; (4) requirements for U.S. person information, including as to purpose, retention, and dissemination, are met; and (5) there are safeguards in place to ensure that the information is used in a manner that satisfies IC standards for "information accuracy, quality, and reliability," including those in ICD 203. Although this guidance is more than a decade old, we believe that it is valuable and could be updated as discussed further below.

3.2. (U) CAI. As discussed above, the definition of PAI in modern IC guidelines includes CAI to the extent that it "is available to the public by subscription or purchase." That description applies to much CAI, and as previously noted it is the focus of our report, to the exclusion of CAI products that are available only to governments. As noted above, IC elements' guidelines recognize that while some CAI is PAI, 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. Approaching the issue from the other side, there are certain legal restrictions on providing CAI to the U.S. government as opposed to other purchasers (see, e.g., 18 U.S.C. § 2702(a)(3)), as well as potential limits on those restrictions (see, e.g., 18 U.S.C. § 2511(2)(F)). In substantial part, however, CAI is available to the IC much as it is to the general public, other private-sector entities and non-governmental organizations (NGOs), and foreign governments.

3.3. (U) CAI Under IC Guidelines. As we understand it, here is the process, in the form of issues and questions, that the CIA and DOD Attorney General guidelines (issued under Section 2.3 of Executive Order 12333) prescribe for potential acquisition and treatment of CAI (other IC

elements have their own guidelines, some of which at this writing are in the process of being revised).

3.3.1. (U) Authorized Purpose. This is required for all activity under the IC guidelines. For example, as a general matter, the CIA may collect information, including information concerning U.S. persons and U.S. Person Identifying Information (USPI or USPII, depending on the agency), only if the collection has "a purpose consistent with [lawful] CIA authorities and responsibilities." CIA Guidelines § 3.3. Similarly, as a general matter the DOD Manual permits intentional acquisition of USPI "only if the information sought is reasonably believed to be necessary for the performance of an authorized intelligence mission or function assigned to the [DOD] Component" conducting the acquisition. DOD Manual § 3.2.c; see *id*. § 3.2.f(2). These baseline requirements preclude, for example, intelligence collection for domestic political purposes. See, e.g., CIA Guidelines § 3.3 ("CIA is not authorized to and shall not collect or maintain information concerning U.S. persons solely for the purpose of monitoring (1) activities protected by the First Amendment or (2) the lawful exercise of other rights secured by the Constitution or laws of the United States … CIA is not authorized to and shall not engage in any intelligence activity, including dissemination of information to the Executive Office of the President, for the purpose of affecting the political process in the United States"); see also The Attorney General's Guidelines for Domestic FBI Operations. These limitations are themselves derived from and consistent with Sections 2.3 and 2.4 of Executive Order 12333 and other provisions of law.

3.3.2. (U) Publicly Available. Where the IC has an authorized intelligence purpose, and the information it seeks is reasonably believed to be necessary for that purpose, it generally may collect information, including USPI, if the information is publicly available. See, e.g., DOD Manual § 3.2.c.(1); CIA Guidelines §§ 4.2(a), 4.21.

3.3.3. (U) Scope of Collection. The CIA's guidelines permit the use of a collection technique "only if a less intrusive technique cannot acquire intelligence of the nature, reliability, and timeliness required," and they treat collection of publicly available CIA as a "basic" collection technique, generally the least intrusive category. CIA Guidelines §§ 4.1-4.2. However, the CIA Guidelines also require (§ 3.3) that in "any collection activity, the CIA shall collect only the amount of information reasonably necessary to support [an authorized] purpose." Where a collection exceeds the agency's ability promptly to evaluate all of the collected information for retention, the CIA guidelines require the approving official to document "the collection technique(s) employed, including any reasonable steps that were or will be taken to limit the information to the smallest separable subset of data containing the information necessary to achieve the purpose of the collection," such as the use of "filters or similar technology" that "should be applied as early as practicable in the course of the collection activity." *Id*. § 5.2(c). The CIA Guidelines explicitly address and require additional documentation for bulk collection (information collected without discriminants). See *id*. §§ 5, 12.2. The DOD Manual provides that in addition to using the "least intrusive means" of collection, § 3.2.(f)(3)(a), "in collecting non-

publicly available USPI," DOD components "will, to the extent practicable, collect no more information than is reasonably necessary." DOD Manual § 3.2.f.(3)(a), 3.2.f.(4). By its terms, this last requirement does not apply to PAI, although DOD's rules on the volume, proportionality, and sensitivity (VPS) of USPI, discussed immediately below, do apply if the CAI includes USPI. Current understandings of CAI as a subset of PAI mean that IC elements are essentially encouraged to acquire CAI, when it is PAI, over other sources of information.

3.3.3.1. (U) Clarification of Current Guidelines. The IC may want to recalibrate, clarify or consider its understanding of whether and how the preference for collection using the "least intrusive means" relates to a preference to collect information that is necessary for an authorized purpose. The increasing availability of CAI means that potentially sensitive information on large numbers of persons may be PAI. In some cases, therefore, collecting a large dataset using commercial means might invade privacy more than a narrower collection using means that target a specific individual or smaller group. New guidance from some IC elements addresses some of these issues (see, e.g., discussion of DIA procedures below).

3.3.4. (U) Volume, Proportion, Sensitivity (VPS) of USPI. The DOD Manual defines a category known as "Special Circumstances Collection" according to the "volume, proportion, and sensitivity of USPI likely to be acquired, and the intrusiveness of the methods used to collect the information," including when the information is PAI. DOD Manual § 3.2.e. When "special circumstances exist, the DOD component head or delegee must determine whether to authorize the collection and, if so, whether enhanced safeguards are appropriate." *Id*. The CIA Guidelines address "volume proportion, and sensitivity" of USPII in "exceptional handling requirements" that apply to unevaluated data sets and might require "additional access approvals or additional training requirements," among other things. CIA Guidelines § 6.2. Some IC elements have established, or are in the process of developing, more detailed VPS guidance, as discussed in Part 4.

3.4. (U) CAI & *Carpenter*. Although to our knowledge the IC has not arrived at a community-wide formal position on the issue, at least one IC element, the Defense Intelligence Agency (DIA), has advised Congress in writing that, as of January 15, 2021, it "does not construe the *Carpenter* decision to require a judicial warrant endorsing purchase or use of commercially-available data for intelligence purposes." We do not express a view on the legal merits of this position, in keeping with our TOR, but in an effort to provide context for readers of this report, we believe that it may rest on one or more of the following theories: (1) certain forms of CAI do not implicate the privacy or Fourth Amendment rights of any data subject (e.g., topographical maps); (2) some CAI involves data and other factors that bring the acquisition outside the scope of *Carpenter*, meaning that normal third-party doctrine (e.g., *United States v. Miller*, 425 U.S. 435 (1976)) extinguishes any rights in the data subject; (3) for data types and acquisition modes that are subject to *Carpenter*, the decision does not apply to the "special need" of intelligence collection conducted by the IC; (4) even if *Carpenter* does apply, it would at most create a shared Fourth Amendment interest among the data seller and the data subject, meaning that

the seller may consent unilaterally to sell the CAI, at least where the subject is not present and objecting to the sale (e.g., *United States v. Matlock*, 415 U.S. 164 (1974); *Fernandez v. California*, 571 U.S. 292 (2014) – a consent-based Fourth Amendment doctrine that is orthogonal to and unchanged by *Carpenter*.

(U~~//FOUO~~) At the same time, however, ODNI has taken the position in writing that, while *Carpenter*'s reach remains uncertain, the IC will collect persistent location data under FISA's provisions requiring probable cause and applicable to collection of communications "contents" rather than metadata, which is a clear effort to hedge against the possible application of *Carpenter* to foreign intelligence collection. This is not meant to suggest internal legal disagreement within the IC, but only to say that IC elements can and have made policy judgments designed to hedge against the possibility that *Carpenter* applies beyond its facts, or otherwise to address the concerns that underlie it.

4. (U) **RECOMMENDATIONS**

(U~~//FOUO)~~ We have three recommendations. ***First****, the IC should develop a multi-layered process to catalog, to the extent feasible, the CAI that IC elements acquire. This will be a complex undertaking requiring attention to procurement contracts, functionally equivalent data acquisition processes, data flows, and data use. The IC cannot understand and improve how it deals with CAI unless and until it knows what it is doing with CAI.* ***Second****, based on that knowledge, the IC should develop a set of adaptable standards and procedures for CAI, governing and requiring regular re-evaluation of acquisition and other decisions.* ***Third****, as part of this set of standards and procedures, and/or as a complement to it, the IC should develop more precise sensitivity and privacy-protecting guidance for CAI. PAI is no longer a good proxy for non-sensitive information; today, much CAI is very sensitive, and the IC therefore needs to develop more refined approaches.*

4.1. (U~~//FOUO)~~ Recommendation #1: The IC Should Learn How It Acquires and Uses CAI. As discussed in Part 1, changes in digital technology and related factors have created a large and growing market for CAI that includes significant VPS of USPI but remains PAI under current IC guidelines. CAI is very valuable as a source of intelligence insight and creates significant risks to privacy. But the IC does not currently have sufficient visibility into its own acquisition and use of CAI across its 18 elements. Accordingly, our first recommendation is for the IC to implement a process that affords it better insight, on a going-forward basis, as to that acquisition and use.

4.1.1. (U~~//FOUO)~~ The Value of Understanding. Given the increasing importance of CAI, and the highly dynamic nature of available offerings and markets, the IC is rightly focused on understanding its own collection and use of CAI. Insight gained from 18 IC elements could inform community-wide best practices in several areas, including means and terms of acquisition, analysis and other approaches to use and exploitation (e.g., increased awareness of the intelligence value of CAI to the IC and to our adversaries), awareness of privacy and other sensitivities, and applicable legal and policy rules and frameworks. Cf. ICD 501. Logically inconsistent approaches to CAI (as opposed to mere differences in approach, which properly may result from differences in mission, authorities, and other factors) can be found and addressed. In addition, overseers will rightly pose questions about the IC's approach to CAI, and the IC should be able to answer those questions with high fidelity and confidence.

4.1.2. (U) Prospective Cataloguing Effort. For three main reasons, we recommend that the IC pursue a forward-looking and recurring effort to understanding its own use of CAI. As part of that process, of course, the IC will need to navigate security and counterintelligence concerns.

(U) First, prior retrospective data calls have not fully succeeded. An attempt from the beginning of 2021 did not return comprehensive and reliable results, and – in part for that reason – the data call underlying our report sought only representative samples of CAI. That data call has served us well, and when combined with insights from our discussions with IC elements we

believe it provides a suitable foundation for our report and recommendations. Our report is not, however, based on anything approaching a complete survey of the use of CAI by the IC, and difficulties in accessing historical information about the use of CAI informs our recommendation for a new, forward-looking approach. Depending on what is revealed by that forward-looking approach, significant new work may be required. For example, as noted throughout this report, our report addresses CAI that is publicly available; if it turns out to be the case that the IC acquires and uses a significant amount of CAI that is not PAI (e.g., because it is sold only to governmental customers, not to the general public), then further analysis on that issue probably would be necessary.

(U) Second, we believe that a prospective effort will be valuable. The IC's acquisition and use of CAI, as well as the overall market for CAI, is very dynamic, making a retrospective survey less informative and useful for developing new approaches.

(U)Third, and relatedly, a forward-looking process that recurs could capture both current and future states of affairs concerning CAI, allowing the IC to keep up with what we expect will be significant developments over time.

4.1.3. (U) Multi-Layered Cataloguing Effort. We recommend that the IC's forward-looking process for cataloguing CAI be multi-layered.

(U) At the outset, the cataloguing effort should include formal contracts and procurement decisions, as well as functionally equivalent agency-specific data acquisition processes, because these will help identify CAI when it first arrives at (or becomes available to) an IC element. By "functionally equivalent agency-specific data acquisition processes," we mean to cover acquisition of CAI that occurs without a formal procurement decision, such as when CAI is provided (for ingestion or via a licensed right of access) to an IC element from a non-IC element, including U.S. Title 10 elements, law enforcement, foreign governments, and non-governmental organizations. These processes may vary from one IC element to another and over time within an IC element. Even within an explicit procurement setting, the role of CAI may not always be apparent – e.g., when a contract is for services, rather than for provisioning of CAI *per se*, but the services in question require the use of CAI by the service-provider.

(U) Given these complexities in the acquisition of CAI, we also recommend that the IC focus detection efforts at later stages of the information lifecycle, including in the process of planning for and actually using data. We assess that this multi-layered approach is the best way efficiently to begin capturing CAI acquisitions and use, including CAI acquired in bulk (or otherwise in substantial amounts).

(U) In particular, key inputs to the cataloguing process may include (1) documentation reflecting the purchase, license, or other acquisition of a CAI dataset; (2) audits by chief information officers (CIOs) and chief data officers (CDOs) responsible for monitoring data flows

across agency systems and repositories; and (3) sourcing, citation and survey data from collection officers and intelligence analysts reflecting the exploitation of specific CAI sources to support tipping, queuing, finished intelligence (FINTEL), and other intelligence products.

(U) If the IC finds that it acquires CAI through mechanisms outside the scope of what we have described in this report, then of course those mechanisms should be examined as well. Here, as in the definition of CAI, attempts to describe a formal scope of effort should not obscure the functional focus on gaining the best possible understanding of the CAI that is actually being acquired and used by the IC.

4.1.4. (U) Common Taxonomy and Understanding. A central goal of the forward-looking process should be to develop an IC-wide common taxonomy and understanding of CAI, to permit meaningful comparisons and analysis at the scale of current and anticipated future CAI operations. Of course, different IC elements will naturally and rightly adopt different standards and procedures for CAI, according to their missions, authorities, need for CAI (and the sensitivity and other attributes of the CAI they need), and other factors. But our assessment is that current practices vary more, and more unsystematically, than is best. Put differently, the IC's approach to CAI so far has been mainly federated, with individual elements operating as what might be called laboratories of CAI governance. Cf. *New State Ice Co. v. Liebmann*, 285 U.S. 262, 311 (1932) (Brandeis, J., dissenting). We think it is now time for the IC to assemble and identify best practices from the range of current practice, addressing both operational and risk/sensitivity frameworks, as discussed in our second and third recommendations, below.

4.2. (U) Recommendation #2: The IC Should Develop a Set of Adaptable Standards and Procedures for CAI. The IC should adopt an end-to-end approach for CAI. The IC does not currently have, and in our view should develop, a set of adaptable standards and procedures for CAI that can be applied across the community. This can be done centrally, for the IC or the Defense Intelligence Enterprise (DIE) as a whole, and/or at individual IC elements (where the approaches could vary from one element to another as long as they are consistent in principle). Either way, as the IC develops approaches to CAI, it will need to keep in mind IC elements' authorities and needs, and the ways in which they approach related issues. There is a well-understood tension, and need to balance, between consistency of approach across IC elements for any single data type, and consistency of approach within an IC element for related data types. We discuss immediately below the main issues of substance, structure, and process that we recommend be included in CAI standards and procedures, and then review and assess examples of such standards and procedures that are currently in effect at certain IC elements. Our third recommendation, discussed further below, addresses the need for enhanced sensitivity guidance for CAI, which overlaps in part with the issues discussed here.

4.2.1. (U) Issues. The IC should develop standards and procedures for CAI that address, among other things, the following issues:

- (U) Mission Analysis to Identify Need/Value. What problem is the acquisition of CAI designed to solve? How important is the problem, and how difficult is it to solve?

- (U) Fit Between Mission and CAI Data Set, Proposed Use. How will the CAI data set solve or address the problem? What other possible approaches are there for addressing the problem or meeting the need?

- (U) Vendor and Data Quality. This includes vendor capacity and longevity as well as quality/reliability issues pertaining to the vendor and its data sources and personnel. These factors may be more applicable to domestic commercial information acquisitions than foreign acquisitions. Can the vendor reliably meet the IC element's needs over time? What assurances of quality are available? Can the vendor adapt if circumstances or needs change?

- (U) Acquisition Mechanics. This includes where, how, and when CAI will be acquired; whether it will be ingested or accessed at the vendor; whether acquisition is overt or covert; whether USPI and other US data will be excluded at the vendor, at initial ingest by the IC element, in query returns, and/or not at all. It also includes ways in which acquisition mechanics may affect the IC's ability to use the data towards the mission need, including through the end user interface and application programming interfaces (APIs) if applicable. The USA Freedom Act illustrates the importance of the policy and engineering issues raised by acquisition mechanics.

- (U) Data Security. This involves potential counter-intelligence risk in the vendor and/or the method of acquisition, and in data storage within the IC element (it overlaps to some degree with the vendor and data quality review described above).

- (U) Sensitivity and Legal Review. This is focused on privacy protection and VPS of USPI and addressed further in Recommendation #3 below.

- (U) Auditing Use of CAI. This involves keeping copies of queries and similar uses of the CAI, and also finding ways to measure actual use and value over time.

- (U) Periodic Re-Evaluation. Finally, there should be a process to reconsider CAI acquisition and other decisions, to avoid inertial automatic renewal of contracts, and to take note when CAI data sources, or the use of CAI within the IC, change materially over time. Cf. ICD 203 and ICD 206, and more general principles of information integrity supporting standards of analytic tradecraft. In Recommendation #3, discussed below, we address re-evaluation to address VPS and sensitivity issues, a similar process with a different purpose than the re-evaluation discussed here.

- (U) Other Structural and Procedural Issues. Some IC elements have a dedicated unit or sub-unit focused on acquisition of CAI. Others have committees or working groups drawn from personnel in relevant sub-units. Both approaches are potentially viable, but IC elements should develop a regular process for addressing all of the issues listed above in regular order and with the benefit of relevant personnel. Results of decisions should be documented and provided to a central authority and assessed together periodically to discern best practices.

4.2.2. (U) Examples of Current CAI Approaches. Several IC elements have established standards and procedures to guide decisions on the acquisition and use of CAI. We review below the approaches taken by three agencies (relevant source materials are in the appendices). Some IC elements also have electronic Data Handling Forms (or the equivalent) governing acquisition of CAI that helpfully standardize aspects of the process for making CAI procurement decisions (or their functional equivalent).

4.2.2.1. (U) Treasury. The Treasury Department has chartered the Office of Terrorism and Financial Intelligence Data Governance Board. We have three main observations on the charter. First, the statement of objective and scope in Part B of the charter, and the statement of the Board's responsibilities in Part C, are general but broad enough to embrace the acquisition and handling of CAI, and it is our understanding that they are applied to CAI. But they do not explicitly apply to such acquisition and handling, and as Treasury advised, the Board was not created to address CAI *per se*, but rather to facilitate data integration and information sharing among all TFI components (one of which, OIA, is a member of the IC).

(U~~//FOUO)~~ Second, while the Board allows for subject-matter experts to participate at the Chair's discretion under Part D.8 of the charter, the Board's regular members do not include representatives from privacy/civil liberties, which means that it may not be well positioned to address issues including VPS. We understand that privacy experts are, at least in some cases, invited to the Board's sessions, and Treasury has advised that privacy considerations do factor into the board's decision.

(U) Third, while the Board has an objective and responsibilities, it does not appear to have any explicit authority over CAI or other matters.

4.2.2.2. (U) Department of Homeland Security (DHS). Like the Treasury Department's Data Governance Board charter, the charter for the DHS Data Access Review Council (DARC) could be adapted explicitly to address issues with CAI, and it is our understanding that DHS currently uses the DARC to review bulk CAI acquisitions.

(U) The DARC's members explicitly include representatives from DHS legal, policy, privacy and civil liberties elements.

(U~~//FOUO)~~ The DARC charter calls for "automatic review" in cases involving an internal or external "bulk data transfer" of PII, and "discretionary review" of other transfers upon the nomination of any DARC member with concurrence of other members, unless the transfer in question has already been approved by higher authority after a review for "legal and policy sufficiency and privacy and civil rights and civil liberties adequacy."

(U~~//FOUO)~~ Highlighting an issue about common use of vocabulary across IC elements, the DARC charter uses "bulk data transfer" to refer to the transfer of "large quantities of intelligence or information, a significant portion of which is not reasonably likely to have any ultimate intelligence or operational value to the recipient." This is similar to how CIA defines and treats "unevaluated data" (see CIA Guidelines §§ 12.22 (definition) and 6.2 (rules)), as discussed above.

4.2.2.3. (U) [redacted] The most mature set of standards and processes governing CAI that we reviewed came from [redacted]

(U) Among the documents we saw, these [redacted] best represent the kind of end-to-end process for CAI that we think is desirable. [redacted]

4.2.3. (U) Assessment of CAI Examples. As the foregoing examples show, there is considerable variation in the approaches to CAI that are currently in effect at IC elements. Some of this variation makes sense in light of varying missions, authorities, and uses for CAI, and much of it is explainable in light of differences in historical experience with CAI. The use of PAI in general, and of CAI in particular, that includes detailed information concerning large numbers of individuals is a relatively new intelligence discipline and still evolving rapidly. We certainly do not mean to say that every IC element must adopt a version of [redacted] very detailed procedures. As noted above, however, we think that IC elements should now come together, review best practices and approaches, and adopt standards that reflect their collective experience, as well as the recommendations in this report. We believe that such an effort will result in more uniform (albeit not identical) approaches to CAI across the IC.

(U) Almost all of the IC elements' acquisition procedures governing CAI that we reviewed are focused on operational and counter-intelligence concerns rather than privacy and sensitivity. To be sure, some of the procedures focus on governance and include legal personnel in decision-making, and some make explicit reference to civil liberties and legal review. But the documents

memorializing approaches to CAI do not address privacy and sensitivity with the same level of rigor and focus that they devote to other issues. That leads to our third and final recommendation.

4.3. (U) Recommendation #3: The IC Should Develop More Precise Sensitivity (VPS) Guidance for CAI. As noted above, CAI can include sensitive information with a high volume, proportion, and sensitivity (VPS) of USPI. Many IC elements' guidelines have provisions that are designed to address VPS concerns in the acquisition, retention, and dissemination of information, including but not limited to CAI. We think those VPS provisions are sound and point in the right direction. As set forth in IC guidelines, however, the VPS provisions are general, in the sense that they afford considerable discretion both on when they apply and how they apply (e.g., what they require to protect privacy). We believe that the IC should develop guidance that refines and applies VPS standards more precisely and explicitly to CAI. Again, the guidance and approach need not be identical at each IC element. We discuss immediately below the main issues of structure, process, and substance that we recommend be included in the guidance, and then review and assess three examples of VPS guidance that are currently in effect (or in development) at certain IC elements. Some of the issues discussed here overlap with our second recommendation, discussed above.

4.3.1. (U) Structural and Procedural Issues. In developing VPS guidance for CAI, IC elements should consider, among other things, the following structural and procedural issues:

- (U) Required involvement of relevant parties at all stages, for the most sensitive cases including legal, privacy, and civil liberties personnel within IC elements.

- (U) VPS assessments generally being made prior to acquisition, or at least prior to analytic use of CAI (with a traditional emergency exception allowing prompt post-acquisition assessing and reporting, with an adequate explanation), ideally integrated or coordinated with CAI acquisition reviews discussed in Recommendation #2 (to avoid VPS concerns being raised as an afterthought or too late in the process).

- (U) Approval requirements, with higher levels of approval required for more sensitive cases, including the possibility of approvals by IC element heads in the most sensitive cases.

- (U) Documentation, retention, and availability to relevant personnel of assessments, approvals, and mitigation measures adopted, in keeping with need-to-know and related security principles, to enhance institutional memory.

- (U) Re-evaluation of VPS assessments and measures, both on a regular basis (e.g., annually) and as circumstances change (e.g., in some cases where material, new information sources are added by the vendor to a purchased CAI data set, or significant

new uses are found for previously collected CAI). In Recommendation #2, discussed above, we address re-evaluation to address mission needs, a similar process with a different purpose than the re-evaluation discussed here.

- (U) Forwarding of assessments and other documentation to ODNI (and/or other central authorities, such as USDI or elsewhere in DOD), and a formal mechanism for periodic review to allow comparisons and discussion of best practices across IC elements, to inform refinements and other development of new guidance, consistent with need-to-know and related security and counterintelligence requirements.

4.3.2. (U) Substantive Issues. Apart from the structural and procedural issues above, we recommend that IC elements also consider the following substantive issues in developing VPS guidance for CAI:

- (U) Sensitivity of the CAI – e.g., concerning protected constitutional rights (including religion, speech, reading, association, and political activities), precise and persistent location, sexual activity, and embarrassment and risk to USPs if CAI is disclosed, in keeping with the discussion of sensitivity in Part 2 of this report. How do general VPS principles apply in these specific contexts?

- (U) Deanonymization/reidentification issues. To what extent are data that have been anonymized less sensitive if the IC element can, without undue difficulty, reverse the anonymization or otherwise identify individuals?

- (U) Importance of mission served by CAI (to balance against sensitivity of CAI). IC elements should conduct an explicit analysis and balance of sensitivity risks and mission benefits.

- (U) Strength of nexus between CAI and mission, and availability, feasibility, costs, and risks of (less intrusive) alternatives.

- (U) Ability to filter USPI prior to ingestion (e.g., at the vendor or through an intermediary, before it is made available for operational or analytic use at an IC element), recognizing that because the USIC is very likely the only consumer of CAI that would want or need to eschew USPI, this may be inconsistent with covert acquisition.

- (U) Traditional minimization approaches and techniques, including ability to acquire CAI via access to data at the vendor rather than ingestion of data in bulk, limits on retention, access, querying, other use, and dissemination of CAI, and possible requirements for special training of relevant personal and auditing of queries and other uses of CAI.

- (U) Availability of other privacy-protective measures in light of the need and anticipated use of CAI (e.g., masking, differential privacy techniques, homomorphic or other forms of encryption) that may not be available or appropriate for all missions and anticipated uses.

4.3.3. (U) Examples of VPS Guidance. Several IC elements have established, or are developing, more refined VPS guidance, including for use with CAI. These efforts also represent a step in the right direction. We review below the approaches taken by three agencies (relevant source materials are in the appendices).

4.3.3.1. (U) DIA. The Defense Intelligence Agency published Procedures for Special Circumstances Collection in DIA Guide 5148.1-2 (February 23, 2021). These procedures provide "guidance to DIA personnel for evaluating whether a collection opportunity should be considered a special circumstances collection." *Id*. § 2. Where a special circumstances collection is found and authorized, "the collecting DIA element in consultation with [the DIA Office of Oversight and Compliance] must also consider whether enhanced safeguards are required to protect access to the information." *Id*. § 4.3. Apart from legal and policy restrictions, the following factors are to be considered (*id*. §§ 4.3.1-4.3.5, sub-section numbering omitted):

- (U) Civil liberties and privacy implications of the collection;

- (U) Potential for substantial harm, embarrassment, inconvenience, or unfairness to U.S. persons if the information is improperly used or disclosed;

- (U) Potential future use of the information being retained and the types of queries or searches expected to be conducted;

- (U) Length of time the information will be retained; and

- (U) Practical and technical difficulties associated with implementing any enhanced safeguards.

(U) If enhanced safeguards are deemed necessary under these factors, one or more of the following measures can be used (*id*. §§ 4.4.1-4.4.5, sub-section numbering omitted):

- (U) Procedures for approval for access to and audit of any searches;

- (U) Procedures to restrict access or dissemination including limiting the number of personnel with access or authority to search; establishing a requirement for higher-level approval or legal review before or after access or search; or requiring higher-level approval or legal review before or after U.S. person information is unmasked or disseminated;

- (U) Use of privacy-enhancing techniques, such as information masking that indicates the existence of U.S. person information without providing the content of the information, until the appropriate approvals are granted;

- (U) Use of access controls, including data segregation, attribute-based access, or other physical or logical access methods;

- (U) Additional protective retention measures or training as required.

(U) Before collection occurs (or as soon as possible after it begins, with an explanation of why collection began before authorization and why continued retention of any previously collected information should be authorized), the "DIA collecting element routes a written summary of the results of its evaluation in paragraphs 4.2 through 4.4 … in a staff package to the appropriate delegated decision authority." *Id*. § 5. The package is then coordinated with the Office of General Counsel and other appropriate DIA elements. *Id*. If and when collection is approved, OOC notifies the DOD Senior Intelligence Oversight Officer. *Id*. § 6.2.

(U) In addition to agency-specific guidance, it is our understanding that DOD is nearing completion of a Department-wide policy on enhanced safeguards which will impose restrictions on the use of certain sensitive forms of CAI.

4.3.3.2. (U~~//FOUO)~~ NSA. The National Security Agency (NSA) adopted NSA/CSS Policy Memorandum 2021-01, Special Circumstances: Guidance for Intelligence Collection of U.S. Person Information, effective for a one-year period beginning March 10, 2021. The NSA memo resembles the DIA Guidance in that it "prescribes the implementation of NSA/CSS procedures for considering whether an intelligence collection opportunity may constitute Special Circumstances Collection requiring enhanced safeguards under paragraph 3.2.e. of" the DOD Manual. NSA Memo ¶ 1. Covered collection opportunities under the NSA memo expressly include those involving information that is "commercially acquired or voluntarily provided," as well as SIGINT, whenever the information in question "is to be retained in a repository for operational purposes." *Id*. ¶¶ 1-2. The NSA memo expressly does not apply to "collection decisions regarding individual foreign intelligence targets," as opposed to, e.g., collection of unevaluated or bulk data; "nor does it apply to analyst queries or disseminations of lawfully collected intelligence information," or to efforts under NSD 42 to secure U.S. government systems (see DOD Manual § 3.1.a.(3)). *Id*. ¶¶ 1-2. The NSA memo explains that it is not a substitute for consultation with NSA lawyers. *Id*. ¶ 3. All approved Special Circumstances collections are to be reported annually to DOD. *Id*. ¶ 10.

(U~~//FOUO) U~~nder the NSA memo, an element of NSA/CSS that is "considering an intelligence collection opportunity" must generally conduct a "Special Circumstances Collection Assessment" (SCCA) "to determine whether [the collection opportunity] includes the

acquisition of USPI that raises special circumstances." *Id*. ¶ 11. In general, NSA's guidance emphasizes measures that prevent such acquisition, by requiring that to "the extent practicable, before collection … organizations will reduce the risk of acquiring USPI that is not responsive to the [mission] purposes of the collection." *Id*. ¶ 18. The NSA memo explains that "post-collection mitigations do not affect" whether special circumstances are found to exist, but "may affect the appropriate decision level for approval of Special Circumstances Collection." *Id*. It is clear that NSA would prefer to filter out unnecessary USPI before collection, a laudable goal.

(U) Where USPI cannot be filtered before collection, an SCCA is generally required with respect to CAI and other non-SIGINT collection opportunities but is expressly not required in four defined situations (*id*. ¶ 13.b.):

> 1) (U) The collection is limited to data or information that is not reasonably anticipated to include USPI, such as statistics or machine-to-machine data (e.g., network infrastructure interactions, netflow, internet routing information);
>
> 2) (U) The collection is limited to data or information that is available to the public at large (e.g., telephone listings, technical journals, newspapers, and books), provided that such collection is not reasonably anticipated to include information concerning USPs resulting from negligence or theft (e.g., hacked or stolen data) and is also not reasonably anticipated to include highly sensitive USPI as further described in paragraph 15;
>
> 3) (U) The collection is provided with consent of an individual or organization in accordance with [the DOD Manual]; or
>
> 4) (U) The collection is not expected to include USPI or is not otherwise governed by [the DOD Manual].

(U) The first two of these four exclusions raise significant questions centered on the application and meaning of the definition of USPI.

[redacted]

[redacted]

(U~~//FOUO)~~ The second exclusion depends largely on how NSA applies the definition of USPI to (publicly available) CAI. The second exclusion generally covers such CAI because it applies to "information that is available to the public at large," including by paid subscription, but it expressly does not include CAI that is "highly sensitive" as defined in paragraph 15. Thus, under Paragraph 13.b. of the NSA memo, an SCCA is required for the collection of "highly sensitive" CAI (that is not subject to the other exclusions).

(U) The definition of "highly sensitive" in paragraph 15 of the NSA memo refers back to the VPS definition of "special circumstances" in the DOD Manual but develops further the meaning of "sensitivity." Paragraph 15 explains that the "sensitivity of USPI" depends on "the potential for substantial harm, embarrassment, inconvenience, or unfairness to any USP if the information is improperly used or disclosed," which is very similar to Section 4.3.2. of the DIA Guidance discussed above, and relevant to some of the concerns raised in Part 2 of this report. Paragraph 15.b.1. of the NSA memo goes on to provide:

> (U) Special circumstances exist if the type of information to be collected relates to or aggregates many data types concerning sensitive activities of any identifiable USP. Sensitive activities include political participation, practice of religion, medical information, membership or participation in organizations or associations, financial data, protected speech, location over time, and protected class demographics. Special circumstances also include publicly available data concerning identifiable USPs that the originator did not intend to be made accessible online or otherwise available to the public (e.g., hacked or stolen data).

(U) Standing alone, this language in Paragraph 15.b.1. appears to mean that an SCCA is required, and that sensitive circumstances exist, when NSA collects (a significant volume of) CAI that is "sensitive," hacked, or stolen USPI (and that is not subject to the other exclusions in paragraph 13.b.).

(U~~//FOUO)~~ Under paragraph 15.b.2., special circumstances "do not exist if the type of information to be collected is limited to USPI that people have chosen to share publicly about themselves, unless such information relates to the sensitive activities of any identifiable USP as further addressed above." As we understand it, NSA does not consider metadata associated

with app downloads or website visits to be data that "people have chosen to share publicly about themselves" within the meaning of this provision. In any case, even if that were not so, paragraph 15.b.2. would require an SCCA, and sensitive circumstances would exist, when NSA collects CAI that includes "sensitive," hacked, or stolen USPI or that relates to an identifiable USP (again assuming the other exclusions in Paragraph 13.b. do not apply). We are not aware of any further guidance from NSA on this question, although as noted above it is our understanding that NSA is currently working to institute additional compliance guidance regarding the handling of publicly available information. We think it may be helpful for such forthcoming guidance to address these issues explicitly.

(U~~//FOUO)~~ Netting out the many layers of guidance in NSA's memo, as we understand it, much turns on whether information, including "sensitive" CAI, is "USPI" (or information that "relates" to "any identifiable USP"). As discussed above, the first exclusion in Paragraph 13.b. turns expressly on whether machine-to-machine data is determined to contain USPI (and again, our understanding from NSA is that it does not). And the second exclusion ultimately turns on a similar question under paragraph 15.b.1.-2. We believe that further guidance is needed on both issues, and as noted above it may be currently in development.

4.3.3.3. (U~~//FOUO)~~ CIA. As of this writing, CIA is in the process of developing principles to govern the acquisition and use of commercial data. [redacted]

[redacted]

(U~~//FOUO~~) [redacted]

(U//~~FOUO~~) [redacted]

(U) 4.3.4. Assessment of VPS Examples and Possible Areas of Future Focus. We appreciate and support the effort reflected in the examples reviewed above from DIA, NSA, and CIA. Although VPS guidance for CAI should follow from VPS guidance in general, we believe that CAI presents a sufficiently significant and growing phenomenon to merit specific guidance, as CIA is currently developing in its principles. We support the processes for assessments and approvals in the guidance from DIA and NSA and believe that CIA should develop a similar approach to ensure the proper application of its CAI principles. Although it has made progress, particularly with respect to bulk (or bulky) collection of CAI, further progress needs to be made in developing visibility into and control of the channels through which CIA acquires CAI.

(U) The single most important point, in our view, is that the existing VPS-CAI guidance should be further developed, ideally with examples illustrating the application of standards to cases. We offer four specific areas, drawn from the longer list above, in which such development would be particularly helpful.

(U) First, [redacted] some IC elements seem to be embracing a relatively binary model, in which CAI is non-sensitive if the government could and/or historically did overtly and lawfully acquire it directly, and sensitive if the government could not or historically did not do so. Cf. Part 2 of this report. In the former category, for example, would be a database of newspaper and magazine articles, while the latter category would include bulk, persistent cell site location information (CSLI), which would normally require a warrant (under *Carpenter*), or [redacted]. This binary model may not satisfactorily classify every possible case involving CAI, but it appears to be at least a good beginning. (As noted above, a third category of CAI, that is not PAI at all because it is available only to governments, is beyond the scope of this report, but worthy of further attention in its own right.) We recommend the IC test and refine the model against the known use cases to develop guidance. Cf. DOD 3115.12 (2010). The basic point is that the qualitative nature of CAI may help determine its sensitivity.

(U) Second, in addition to qualitative differences in CAI, quantitative differences are also relevant. CAI that is acquired in bulk will almost always be more sensitive than CAI in smaller data sets. Where bulk acquisition can be avoided, and the volume of acquired data is reduced, it is generally helpful both for intelligence purposes and for the protection of privacy and civil liberties, and it may simplify CAI procedures.

(U) Third, subject to the policy concerns underlying PPD-28's approach to SIGINT, U.S. intelligence law and policy emphasize the protection of USPI, and approaches to CAI should be developed consistent with that emphasis. In general, foreign CAI data sets concerning foreign persons and entities may raise fewer, or at least different, concerns than analogous data sets focused on the United States and/or U.S. persons.

[redacted] Fourth, as discussed with respect to the NSA principles, the IC should develop guidance on how the definition of USPI, and the definition of information that pertains to a known USP, apply in the context of CTD and other CAI. If the *New York Times* can easily de-anonymize persistent location data on U.S. persons, and similar efforts are possible and/or may be undertaken by the IC for AdTech and other CAI, is the information therefore USPI? The question arises, and the guidance is needed, because the term USPI is defined in IC policies as "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," with a recognition that the definition as applied "in a particular context may require a case-by-case assessment by a trained intelligence professional." CIA Guidelines § 12.25; DOD Manual 5240.01 §§ 3.2, G-2. We noted inconsistencies between how different IC elements define and treat USPI, with some treating data as non-USPI because they did not possess other data sets that could be used to reidentify (deanonymize) or because they did not intend to reidentify the individuals in the data. This strikes us as unacceptably narrow; at a minimum, the issue of readily-available deanonymization should be considered closely and more precise guidance provided in the context of CAI.

(U) Beyond these areas, we think that the IC also should at least begin working on assessing and developing more specific guidance for various forms of existing or emerging CAI, including from social media, biometrics, augmented reality/virtual reality (AR/VR), and the Internet of Things. (U) Even if designed for specific areas, of course, this guidance should be consistent with the principles discussed above.

(U//~~FOUO~~) We are agnostic as to whether the guidance should be set out in a stand-alone document devoted to VPS issues in CAI [redacted] added to broader CAI processes (of the sort discussed in our second recommendation above), added as an amendment to existing procedures governing intelligence activities of IC elements, or included in guidance addressing VPS concerns in general (not limited to CAI). The main point is that the guidance be sufficiently clear and specific.

5. (U) **CONCLUSION**

(U~~//FOUO~~) We tried, in Part 1 of this report, to describe CAI for those who are not already familiar with it. Part 1 therefore included a working definition of CAI (and an explanation of why it is important to define); a list of the main sellers of CAI and a brief description of the types of information they make available; an effort to trace the origins and evolution of CAI in the rise of digital data; and a review of how "anonymized" CAI can be reidentified and linked to individuals.

(U) Part 2 explained why the DNI was right to commission our report. It described how CAI can provide intelligence value and identified several examples of IC contracts for CAI. It also addressed non-analytic uses of CAI and counter-intelligence risks in CAI, and the risks that CAI presents for privacy and civil liberties. As we observed at the end of Part 2, CAI is a relatively new, rapidly growing, and increasingly significant part of the information environment in which the IC must function, deserving of focused attention.

(U) Part 3 reviewed in detail the current IC policy and regulatory framework governing CAI, under which, as information that is available to the general public, it is treated as PAI. Part 3 tried to describe the current state of CAI regulation as a baseline for our recommendations.

(U) Part 4 set out our three main recommendations:

(U) First, the IC should develop a forward-looking and recurring process to catalog the acquisition and use of CAI across its 18 elements. The IC cannot understand and improve how it deals with CAI unless and until it knows what it is doing with CAI.

(U) Second, based on the knowledge gained from that process, the IC should develop a set of adaptable standards and procedures for CAI, governing and requiring regular re-evaluation of acquisition and other decisions. We offer several elements that can be included in those standards and procedures, but also recognize that they will need to be adapted for different IC elements with different CAI missions.

(U) Third, as part of this set of policies and procedures, and/or as a complement to it, the IC should develop more precise sensitivity and privacy-protecting guidance for CAI. Again, we offer several suggestions for the development of such guidance.

(U) If some or all of these recommendations are agreeable, the IC will need a mechanism for putting them into effect, and for making any other changes suggested by continued attention to CAI. One possibility, which we believe is worth considering, would be a traditional working group of IC senior officials. This group would be charged with implementing our recommendations (to the extent approved), sharing best practices, collecting and assessing additional information about the IC's use of CAI, and recommending additional improvements

over time. The group might decide to proceed within the framework of our three recommendations, or it might adopt their substance within a different framework. For example, the working group might focus on developing (1) principles, such as utility, privacy, and quality of data; (2) tools and procedures for implementing those principles, such as technological methods for filtering and limiting data before its ingestion or use; and (3) processes and approval requirements for applying those tools and procedures. As noted above, CAI is both increasingly powerful for intelligence and increasingly sensitive for individual privacy, and while we hope that our 90-day report provides a helpful foundation for developing more refined approaches, we believe that continued efforts will be necessary. We appreciate the opportunity to be of service

6. (U) APPENDICES

6.1. (U) Letter and Terms of Reference

6.2. (U) IC Elements' Materials Governing CAI

6.3. (U) IC Elements' Materials on VPS and/or CAI Collection