OFFICE OF THE DIRECTOR OF NATIONAL INTELLIGENCE

2018 Data Mining Report

For the Period January 1, 2018 through December 31, 2018

July 2019

Office of the Director of National Intelligence
2018 Data Mining Report
January 1, 2018 through December 31, 2018

I. Introduction

The Office of the Director of National Intelligence (ODNI) provides this report pursuant to the Federal Agency Data Mining Reporting Act of 2007, section 804 of Public Law 110-53 (codified at Title 42 United States Code section 2000ee-3) (the "Data Mining Reporting Act" or the "Act").

A. Scope

This report covers the activities of all ODNI components from January 1, 2018 through December 31, 2018. Consistent with the Act, constituent elements of the Intelligence Community (IC) will report their activities to Congress through their own departments or agencies.

B. Reporting Requirement

The Act requires that, each year, "the head of each department or agency of the Federal Government that is engaged in an activity to use or develop data mining shall submit a report to Congress on all such activities of the department or agency."[1] The Act defines "data mining" as:

"… a program involving pattern-based queries, searches or other analyses of one or more electronic databases, where —

1) a department or agency of the Federal Government, or a non-Federal entity acting on behalf of the Federal Government, is conducting the queries, searches, or other analyses to discover or locate a predictive pattern or anomaly indicative of terrorist or criminal activity on the part of any individual or individuals;

2) the queries, searches, or other analyses are not subject-based and do not use personal identifiers of a specific individual, or inputs associated with a specific individual or group of individuals, to retrieve information from the database or databases;[2] and

[1] 42 U.S.C. § 2000ee-3(c)(1).
[2] As stated in prior reports, certain analytic tools and techniques, such as link-analysis tools, rely on "personal identifiers of a specific individual, or inputs associated with a specific individual or group of individuals," such as a known or suspected terrorist, or other subject of foreign intelligence interest, and use various methods to uncover links or relationships between the known subject and potential associates or other persons with whom that subject has a

3) The purpose of the queries, searches, or other analyses is not solely — (i) the detection of fraud, waste, or abuse in a Government agency or program; or (ii) the security of a Government computer system.[3]"

C. Report Content

In recent years, we have followed a format that we believe enhances clarity and readability. Specifically, Part II of the report describes activities that meet the definition of "data mining" under the Act, as well as programs that meet some, but not all, of the criteria defining "data mining." We report the latter category of activities in the interest of transparency. Part III provides updates on programs included in the prior year's report. Part IV of this report provides an overview of the Privacy and Civil Liberties infrastructure within which ODNI conducts its activities.

II. New Activities

The ODNI did not undertake any new reportable activities during the reporting period.

III. Updates on Previously Reported Activities

As previously discussed, this section provides updates on programs described in last year's report.

A. NCTC

NCTC continues to conduct "threat analyses" as described in the 2013 Data Mining Report. As noted in the 2013 report, this is an analytic technique to narrow the pool of information within NCTC databases that analysts will assess in response to specific threat reports. This technique does not meet all of the statutorily defined criteria for data mining, but has been reported in the interest of transparency.

B. IARPA

IARPA continues to invest in high-risk, high-payoff research programs that have the potential to provide the United States with an overwhelming intelligence advantage over future adversaries. As a scientific research funding organization, IARPA does not use, nor does it expect to make use of, data mining technology. IARPA programs are experimental by nature, and are designed to produce new capabilities, such as those described in previous reports and recapped here.

"link" (a contact or relationship). Such tools and techniques are not considered to meet the "data mining" definition of the Act.

[3] 42 U.S.C. § 2000ee-3(b)(1).

(i) Cyber-attack Automated Unconventional Sensor Environment (CAUSE) Research Program.

The CAUSE Research Program began in 2016 and is expected to end in 2019. The program is developing a fully automated system to forecast and provide early warnings of cyberattack events (e.g., denial of service, spearphishing, malware installation, and accessing malicious websites) with enough lead-time to effectively protect information systems. Research will identify and validate leading signals from large volumes of both traditional internal sources, such as host and technical sensor data, and unconventional data sources, including help desk ticketing, social media, and the dark web. Research is focused on external sources of publicly available data obtained legally through agreements, free access, subscription, or purchase. CAUSE Research Teams will analyze cyber-attacks and related activities against a particular U.S. business sector/industry (e.g., financial services, energy, defense) to develop forecasting models and cyber entity and event extraction methods that will identify and detect patterns of activities that precede such cyber events of interest. Such activities may contain information about the "attackers"—i.e., the sources of a cyber-attack, and the victim or target of a cyber-attack, which may include an organization or a person. While this information is publicly available, the CAUSE Research Program is not interested in further identifying individuals responsible for the cyber-attack. The CAUSE Research Program does not generate individual cyber attacker's identities and therefore does not constitute data mining as defined in the stature, but has been reported in the interest of transparency. The CAUSE Research Program to date has demonstrated the utility of various types of data features for forecasting cyber events, particularly from the deep and dark web, and continues to research novel data streams and algorithms.

(ii) Deep Intermodal Video Analytics (DIVA) Research Program

The DIVA Research Program began in 2017 and is expected to continue into 2021. The program is developing methods to watch streaming video from multiple cameras and automatically detect activities of interest (e.g., a person carrying something heavy, a car making a U-turn, a person abandoning a package, etc.). The program does not research methods for identifying individuals (e.g., face recognition). Instead, it is focused on detection of activities that could lead to threats (e.g., attacks on a government facility). Video data for a variety of activities was collected in a fully-controlled setting using hired actors under conditions reviewed and approved by an institutional review board (IRB) to ensure compliance with all relevant privacy concerns. Additional data is being collected under conditions reviewed and approved by an IRB to ensure there is no personal identifiable content present in the collections. Research teams are developing algorithms to detect activities in extended videos that contain large segments without activities of interest, typical of security camera footage. During 2018, these algorithms have been evaluated on existing datasets approved by an IRB. The National Institute of Standards and Technology (NIST) is performing the official evaluations under the program. The program to date has demonstrated the ability to recognize and localize activities in video at levels that surpass previous State-of-the-Art. The program continues to research novel

approaches to the activity detection problem to reach false-alarm rates low enough for operational use.

(iii) Mercury Research Program

Mercury developed and tested methods to provide early warning of significant societal events through continuous, real-time regional analysis of diverse SIGINT data. Throughout the program, IARPA performer research teams worked with operational National Security Agency / Central Security Service cloud information systems using commercial off-the-shelf or government off-the-shelf technologies. The performers tested the performance of their prediction algorithms using open source data provided by IARPA as well as performer provided data from the Gallup organization. During this final period, MERCURY had some limited success with respect to population or group behaviors related to military action, civil unrest and terrorist activities. The Mercury program concluded in 2018 and will not be reported in 2019.

C. ODNI Science and Technology (S&T)

(i) Community Video Analytics Lab (CVAL)

The CVAL is an activity led by the IC Video Collaboration Initiative (VCI), under ODNI/S&T. The VCI is a collaborative group of technical domain experts that serves to advance video analytic capabilities across the IC. The CVAL provides technical evaluations of video analytics capabilities and frameworks and manages a repository of test data to support these evaluations. While neither the VCI nor CVAL constitute data mining, resulting advancements in video analytic technologies and capabilities across the community could potentially be applied to support data mining for the purpose of identifying terrorist or criminal actors in video records. In 2018, the CVAL transitioned to the National Geospatial-Intelligence Agency and will not be reported in 2019.

IV. Protection of Privacy and Civil Liberties

The ODNI Office of Civil Liberties, Privacy, and Transparency (CLPT) works closely with the ODNI Office of General Counsel, ODNI components and the IC elements to ensure that appropriate legal, privacy, and civil liberties safeguards are incorporated into policies, processes and procedures that support the intelligence mission. CLPT is led by the Civil Liberties Protection Officer (CLPO), a position established by the Intelligence Reform and Terrorism Prevention Act of 2004. The duties of the CLPO are set forth in that Act, and include: "ensuring that the protection of civil liberties and privacy is appropriately incorporated in the policies of the ODNI and the IC; overseeing compliance by the ODNI with legal requirements relating to civil liberties and privacy; reviewing complaints about potential abuses of privacy and civil liberties in ODNI programs and activities; and ensuring that technologies sustain, and do not erode, privacy protections relating to the use, collection, and disclosure of personal information."[4] In addition, the CLPO also serves as the ODNI's Chief Transparency Officer and

[5] National Security Act of 1947, as amended by the Intelligence Reform and Terrorism Prevention Act of 2004 (IRTPA), 50 U.S.C. § 403-3d.

chairs the Intelligence Transparency Council, and leads and coordinates the IC's efforts to enhance transparency. Before any innovative capabilities or technology could be used in an operational setting, the use of the capability or technology would need to be examined pursuant to Executive Order (EO) 12333, the Privacy Act, and other applicable requirements to determine how these tools could be used consistent with the framework for protecting U.S. persons (USP) information.

The IC has in place a protective infrastructure built in principal part on a core set of USP rules derived from EO 12333. This EO requires each IC element to maintain procedures, approved by the Attorney General, governing the collection, retention and dissemination of USP information. These procedures limit the type of information that may be collected, retained or disseminated to the categories listed in part 2.3 of the EO. Each IC element's Attorney General-approved USP guidance is interpreted, applied, and overseen by that element's Office of General Counsel, Office of Inspector General, and other compliance offices as appropriate. Violations are reported to the Intelligence Oversight Board of the President's Intelligence Advisory Board. In addition to EO 12333, IC elements are subject to the requirements of the Privacy Act, which protects information about U.S. citizens and permanent resident aliens that a government agency maintains and retrieves by name or unique identifier. The IC also conforms to policies and procedures under Presidential Policy Directive (PPD) 28, relating to protections for all personal information contained in SIGINT.

The IC's privacy and civil liberties protective infrastructure is bolstered further by guidance and directives issued by the Office of Management and Budget pertaining to the protection of personally identifiable information and the development, procurement and operation of information technology systems that administer personally identifiable information.

Finally, the IC has developed and established two sets of principles that have been adopted as "foundational" to the IC mission: The Principles of Professional Ethics for the Intelligence Community, and the Principles of Intelligence Transparency for the Intelligence Community. These two sets of principles inform the IC's approaches to applying appropriate protections for the types of activities described in this report.