Technology for Privacy Assurance

J.C. Smart

1 Introduction

Two pillars of a democratic society—Security and Liberty—are challenged by the post-9/11 world: How can an open democracy sustain the former without infringing on the latter? In our new “Big Data” era, a government’s ability to collect, process, analyze, and share volumes of information is commonly regarded as central to its national security and its public safety. But these needs, driven by a desire to detect threats and reduce risk to the aggregate population increasingly have been placed in conflict with the constitutional protections of individual liberties.

Current public opinion often frames this tension as a tradeoff, balancing the sacrifice of some liberties against real or perceived gains in security and safety (Center for Strategic and International Studies 2014; Gilmore 2014; Campos 2014). A decade and a half later, no end to this debate is in sight. But the presentation here posits that security/safety and liberty are not mutually exclusive. Rather, it advocates a paradigm that enables both to be achieved simultaneously, through the careful application of policy and modern technology (Smart 2011). This concept and the prescribed implementation approach is referred here as Privacy Assurance.

2 Information Sharing

The sharing of information across legal and jurisdictional boundaries enables new analytic opportunities. From a national security perspective, witness how the 9/11-hijackers were not only connected via airline data and other transactional records, but in at least two cases by threat information already maintained by the

J.C. Smart (✉)
Georgetown University, Washington, D.C., USA
e-mail: smart@georegtown.edu

J. Collmann and S.A. Matei (eds.), *Ethical Reasoning in Big Data*,
Computational Social Sciences, DOI 10.1007/978-3-319-28422-4_8

U.S. Intelligence Community. In the public safety context, HIV spreads between individuals who increasingly receive care and treatment across many jurisdictional boundaries that span where they live, work, and socialize. The new spectrum of contemporary analytic techniques is often popularized as “connecting the dots.” But localized information “stovepipes” maintained by individual organizations often are not sufficiently rich in their content to discern the complex network of associations and connections across multiple jurisdictions that realistically describe contemporary threats or societal risks. In contrast, such patterns often are quickly revealed when these otherwise disparate information sources can be merged and analyzed in aggregate.

Unfortunately, the merging of information sources can quickly exceed the respective policies and authorities of participating organizations, creating the new tensions to individual liberties and personal privacy. Alternatively stated, while it often may be in the best interests of single organizations spanning various legal and jurisdictional boundaries to share information, there may not be adequate trust among the participants, or authority from the citizenry under whom they serve, to allow such sharing. This reluctance or mistrust can arise from the fear of misuse with insufficient oversight, fear of the exposure of sensitive information, sources, and methods, or the increased risk of unintentional exposure. Trust and fear issues aside, privacy policy in the United States today mandates data minimization—to wit, that civilian agencies should only collect personally identifying information (PII) that is directly relevant and necessary to accomplish the specified purpose of its collection; only retain PII for as long as is necessary to fulfill the specified purpose; and only share data with other agencies when compatible with the purpose for which it was collected. Moreover, U.S. citizens are afforded constitutional assurance to be “secure in their persons, houses, papers, and effects, against unreasonable searches.” Is it possible to achieve national security and public safety goals without eroding such fundamental privacy rights?

The paradigm advocated here takes the Fourth Amendment to the United States Constitution as a basic system requirement. Within this framework from a national security perspective, U.S. law defines “reasonable suspicion” as the standard of law, based on specific and articulable facts and inferences, under which a person may be regarded as being engaged in criminal activities, having been engaged in such activity, or about to be engaged in it. An analog can be readily devised for the public safety sector with “reasonable concern” as the rubric, based on specific and articulable facts and inferences, under which a person may be regarded as being engaged, having been engaged, or about to be engaged in behavior that exposes the public to undue risk.

Reasonable suspicion is the basis for investigatory stops by the police and requires less evidence than probable cause, the legal requirement for arrests and warrants. Analogously, reasonable concern is a basis for required public health organization reporting (e.g. detection of an highly infectious disease) versus higher thresholds requiring quarantine, mandatory evacuation, imposition of marshal law, etc. Reasonable suspicion or reasonable concern are evaluated using the “reasonable person” standard, in which an official (e.g. police officer or public health officer) in the same circumstances could reasonably believe a person has been, is, or is about to be engaged in an activity that seriously jeopardizes the public's security and/or safety.

Such suspicion or concern cannot simply be based on a hunch. A combination of particular facts, even if each is individually innocuous, can form the reasonable suspicion or reasonable concern. This is pivotal to Constitutional law enforcement and to the method for assuring privacy that is laid out below. It describes how reasonable suspicion (concern) can be ascertained from multiple information sources without resorting to unreasonable search. Unreasonable search is interpreted here as any type of investigative process that would reveal information that a reasonable person would regard as private, prior to the establishment of reasonable suspicion/concern or probable cause—and thus protected.

3 Privacy Assurance

So how can reasonable suspicion (concern) be responsibly ascertained from multiple information sources without resorting to unreasonable search, and thus jeopardizing individual privacy? One approach commonly attempted today is the use of anonymization. That is, all discerning PII is removed, sometimes replaced with statistical results versus actual data, sharing only information that is non-identifiable. Unfortunately, in the new "Big Data" era, true anonymization becomes increasingly difficult at increasing scale, as relationships previously hidden among the enormous data complexity can be revealed as processing of larger and larger data volumes from greater numbers of sources continues to grow. Alternatively, anonymization techniques that truly are effective at scale often dramatically reduce the value of the information being exchanged and its ability to enable actionable outcomes. This is particularly apparent in public health applications where the goals are ultimately to genuinely improve the condition of individuals, versus simply a statistical awareness of an aggregate population's inevitable plight.

The privacy approach advocated here posits the existence of a "Black Box." In this context, a Black Box is a physical (or logical) device whose contents are beyond reach: that is, its contents can *never* be examined. The device is specifically engineered so that the information it is fed cannot be revealed to anyone under any circumstances, regardless of authorization, executive privilege, court order, vandalism, or deliberate attack. Information can flow into the Black Box, but once it resides within its boundaries, it can never be accessed. For all practical purposes, the Black Box is considered an impenetrable information container.

Total impenetrability, however, implies a theoretical extreme that likely would be difficult to achieve, or even more important, to verify or accept in the negative. Consequently, this paper treats impenetrability as the condition in which there exist no known exploitable vulnerabilities that would enable access to the contents of the Black Box. While vulnerabilities may exist, an impenetrable Black Box is one

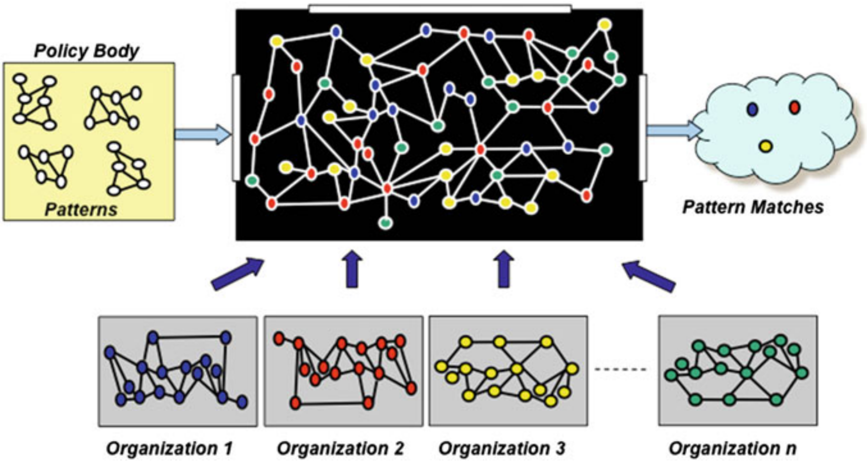


Fig. 1 The privacy assurance "Black Box"

about which a group of reasonable, qualified technical experts will testify that any vulnerabilities inherent in the device's design have been mitigated, using reasonable techniques to assure its security to within a degree of probability asserted as reasonable by a community of such experts.

But what good is a Black Box? Assuming the existence of such a device, it makes possible the ability to "share" private information in unique and powerful ways. However, a new paradigm that governs the notion of analysis and how it can be performed is required.

Figure 1 illustrates the basic privacy assurance concept. At the top center of the diagram is the "Black Box" construct. Across the bottom are representations of independent organizations that span multiple legal and/or jurisdictional boundaries. Each of these organizations via their respective legal charters is authorized to maintain a specific body of information, represented by the colored "dot" networks depicted within each. These information "dots" are connected via "links" that represent relationships that the organization has discerned and maintains, consistent with its legal authorization.

The legal charter of each organization may limit its ability to access or share information and thereby identify the corresponding relationships across the established boundaries. Sharing this information across such a boundary could in fact constitute a breach of law or, alternatively, a breach of public, legislative trust or acceptance. Nevertheless, if such organizations were actually able to share their information, new relationships within the information could be identified from analysis. New patterns of suspicious activity that might impact national security/public safety could be identified and acted upon. This information would constitute "actionable intelligence."

The solution offered here involves placing relevant information from each contributing organization inside of the Black Box. *Information can then be* *connected and processed within, but only without the possibly of human examination or disclosure*. The internal methods used to do the processing are established in contemporary analytic tradecraft. Techniques such as graph analysis and statistical correlation can discover otherwise hidden relationships among billions of data elements. But if such a Black Box is designed to be "non-queryable" by any means, how then can it be of any value?

To address the utility question, the Black Box also has exactly one additional input (on the left in Fig. 1) and exactly one and only one output (located on the right). At the left interface, patterns of specific interest are input to the box. These patterns are template-like encodings of generic information relationships that a duly authorized policy body has reviewed and approved for submission into the box. Put another way, the patterns are a set of analytical rules that define the Black Box's reasonable search behavior. The only patterns that are admissible to the Black Box are those that the policy body has reviewed and has unanimously confirmed as meeting a certain threshold. In this case, the threshold is the set of observable conditions within the Black Box that meet the legal standard for reasonable suspicion or reasonable concern.

Within the Black Box, in addition to the information that it receives from each contributing organization, and the patterns it receives from the policy body, is an algorithm that continuously observes for conditions that match any of the submitted patterns. Upon detecting such a pattern, the Black Box outputs an identifier for the pattern and a set of identifiers for the information that triggered the pattern's detection. This is a continuous process. It is executed in real-time without human intervention, again leveraging current analytic tradecraft. Upon such a detection event, the Black Box would notify the appropriate contributing organizations of the particular identifiers, but without revealing any of the private information it holds within. These organizations could then investigate further, using their existing analytic capacities and legal authorization structures. If permissible by policy and law, additional information could accompany the output notification to expedite investigation. The specification for such auxiliary output information is incorporated into the original pattern definition, enabling the policy body to review and approve in advance, and ensuring privacy compliance throughout.

Output generated by the Black Box would be available to the policy body or alternatively, to a duly constituted oversight body to continuously verify compliance. In other words, while considerable information is flowing into the Black Box, the only aspect that would ever have external visibility is its reasonable suspicion/concern output. This output would be expressed in terms of identifiers that only have meaning to the submitting organization. In this manner, organizations and the citizenry they serve can receive the benefits or information sharing, but without exposing this information to misuse or the risk of privacy invasion in the process.

Under this paradigm, the only information that can be submitted to the Black Box is information that a participating organization has already been authorized to possess (i.e. this process does not address the sharing and analysis of illegally obtained information). Similarly, the only information that is ever outputted from the Black Box is that which has been deemed *in advance* to constitute reasonable suspicion/concern and to meet the standards of law and public policy for protecting individual privacy.

4 Privacy Certification Levels

This work recognizes that the level of privacy assurance obtainable is directly related to the degree at which privacy device "impenetrability" can be achieved, involving a risk–cost benefit tradeoff. Depending upon the nature of the information to be protected, not all information sharing and analysis applications will require the same degree of rigor to ensure adequate privacy protections. For example, the transmission of personal medical information would presumably have a substantially higher level of privacy concern over, say, sharing of publically available property records. Consequently, a multi-level privacy certification rating is envisioned. Analogous with U.S. cryptographic systems (Committee on National Security Systems 2010), the following four levels of privacy certification are proposed:

- *Type 1 Privacy*: a device or system that is certified for national/international governmental use to securely share and analyze private information consistent with the highest level of protections awarded by law and treaty. Type 1 is used to protect information that would result in exceptionally grave damage if disclosed. Achievement of this rating implies that all components of the end-to-end system have been subjected to strict verification procedures, are protected against tampering and subject to strict supply chain controls with continuous oversight.
- *Type 2 Privacy*: a device or system that is certified for governmental and commercial use to securely share and analyze personal information consistent with high levels of protections in conformance with jurisdictional policies and procedures and commercial law. Type 2 is used to protect information that would result in serious privacy damage if disclosed. Achievement of this rating implies that all interface components of the system have been subjected to strict verification and supply chain controls and that all other components have been subjected to reasonable best industry practices for operation verification and supply chain control and oversight.
- *Type 3 Privacy*: a device or system that is certified for public use to securely share and analyze sensitive information. Type 3 is used to protect information that would result in privacy damage if disclosed. Achievement of this rating implies that all components of the system have been subjected to reasonable best industry practices for operation verification and supply chain control.
- *Type 4 Privacy*: a device or system that is registered for information sharing and analysis, but not certified for privacy protection. No assumptions regarding component verification or supply chain controls are made about systems at this privacy protection level.

At a general level, Type 3 systems are composed of components that are designed and integrated using best industry practice. To achieve a higher assurance rating, best industry practice is not considered adequate. For a Type 2 system, while internal components may be commercial items, all interface components must be subject to a rigorous verification process to ensure the validity of all transactions that cross the Black Box boundary. For a Type 1 system, this same rigor must be applied to the entire system, including the design and implementation of internal components and their procurement supply chain. The primary differentiators between these levels ultimately translate to cost. That is, Type 1 systems will generally be more expensive than Type 2 systems, which in turn will be more costly than Type 3 systems, etc. These cost differences are warranted in order to gain higher assurances of privacy protection due to the varying risks associated with the intended applications at each level.

5 Privacy "Black Box" Design

The generic design of a Black Box is shown in Fig. 2. All information that flows into and/or out of the box must pass through carefully designed interfaces that isolate the Black Box internals from the external environment. External data sources at the left side of Fig. 2 are connected to the box via a set of input isolators. These isolators allow correctly encrypted data to flow into the box only from organizations that are properly authorized and authenticated. These isolators enforce a strict one-way flow of data providing no means of internal access or

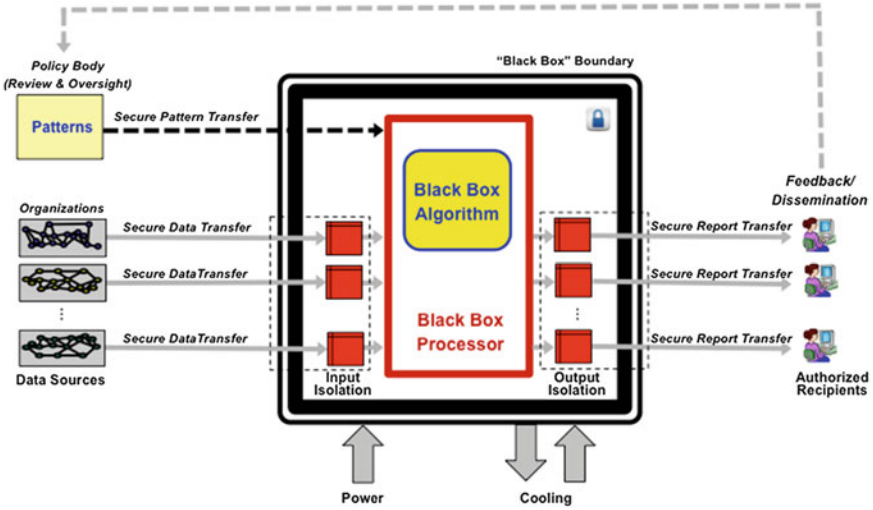


Fig. 2 Generic Black Box design

visibility to any data, status, operating conditions or parameters of the Black Box contents within. While the incoming data received from an organization may well contain private information, accompanying this information for each individual is an identifier (e.g. a unique number sequence) that is assigned by that organization. When a pattern is detected by the Black Box that involves an individual, only the respective identifier for this individual is referenced in the output report. In this manner, no personal information ever leaves the Black Box boundary once it enters.

Within the Black Box is a computer processor that runs an algorithm whose operation is strictly defined by the patterns approved by the policy body. These patterns encode reasonable suspicion/concern policy statements that define the algorithm's behavior. The configuration of this algorithm and its execution of the patterns is carefully controlled and monitored by the policy body to ensure that the Black Box behaves only as they have unanimously specified.

On the right side of Fig. 2 are output isolators that ensure that all output reports that are generated by the internal algorithm flow only to the correct, authorized recipients and that no private information is exposed. The output reports reference individuals via the unique identifiers known and provided by the source organizations. Contained in the reports are indicators of the patterns that the Black Box detected. The policy body controls the specification of these indicators as part of the pattern review and approval process. Unanimous agreement of these indicators is required in advance of the Black Box performing any data analysis. An output feedback loop to the policy body is shown in Fig. 2 for oversight and compliance.

The key aspect of this design is that regardless of what information might flow into the box, the only information that can ever exit is that which was approved and authorized by the policy body as meeting the patterns they have unanimously deemed reasonable. Furthermore, the box itself is implemented in such a manner that these protections cannot be circumvented via tampering. Hence, the implementation cannot provide any back doors, overrides, special authorizations, nor expose any inherent exploitable vulnerabilities, within the limits of the verification techniques and certification process used to specify, design, and engineer its correct operation, commensurate with the assurance level.

6 Example Use Case: Identity Name Resolution

In today's information age, organizations frequently provide overlapping services to individuals. Such overlap can be costly, resulting in unnecessary duplication and expenditure of resources. Resolving this overlap, however, can be extremely complex and time-consuming. Where individuals live, where they work, and how and where they receive these services, and how and when these might change can all greatly vary. Further complicating this process is the incompleteness, errors, and ambiguity in the data that each organization may associate with an individual. The spelling of names, accuracy of birthdates, absence of a consistent universal identifier (e.g. in the U.S., a Social Security Number), etc. all compounds this resolution complexity. Given the sensitive nature of personal information and the complex policies and laws regarding its proper handling, organizations unfortunately are often forced to resort to costly, time-consuming manual methods to identify and resolve discrepancies.

6.1 The Black Box Pilot System

In March of 2015, the first formal application of the Black Box technology was successfully deployed to automate this process in near real-time fashion. The deployment involved three public health organizations working to prevent the spread of HIV within and across their jurisdictional boundaries. Each of these jurisdictions maintains sensitive databases about individuals infected with this disease for their areas. These databases are populated as a result of mandatory reporting procedures followed by the health care providers operating within each of the respective boundaries. To mitigate the spread, it is important that jurisdictions communicate with their neighbors to ensure that individuals remain in care, continuing to receive treatment to help keep their HIV viral counts sufficiently low. As individuals live, work, and receive health care services at varying locations throughout these jurisdictions, resolving identities across the databases has often been a painstakingly slow and difficult process, heightened by the sensitivity of the condition and the importance of protecting each individual's privacy. For this pilot activity, a Type 3 privacy assurance level system was configured. Figure 3 contains an overview of the system's design.

The pilot system consisted of a single, self-contained computer that was physically mounted within a steel reinforced enclosure with multiple security locks (one for each participating jurisdiction). This unit was housed in a non-descript, limited access Tier 3 data center facility managed by Georgetown University with continuous 24/7 video motion detected alarm surveillance. The enclosure was configured such that the computer within could not be removed without resulting in loss of its electrical power. The computer itself was delivered sealed from the factory and was installed and configured only in the presence of security representatives from each organization. The computer was equipped with the most minimal of services, with nearly all external features disabled including the removal of keyboard and mouse input, video display, and unnecessary operating systems components. The disk contents were secured with high-grade encryption. All wireless interfaces (e.g. WiFi and Bluetooth) were disabled, and no external I/O devices were attached nor were ports accessible once secured within the locked enclosure.

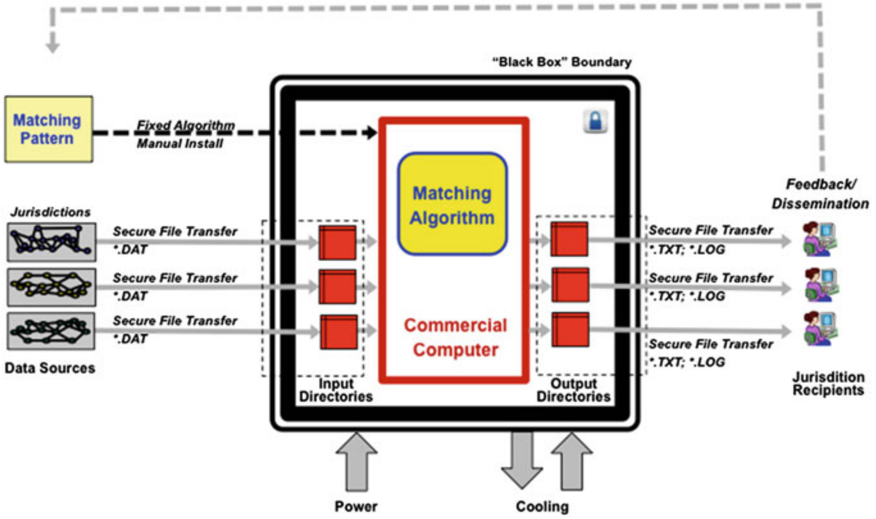


Fig. 3 Pilot (Type 3) Black Box system design

The operating system, network, and supporting firewall infrastructure were configured to allow only secure file transfer access into and out of specific fixed directories, with one directory allocated for each participating organization. The only operations that were permitted by an organization were reading, writing and deleting files in their respective assigned directories. All file accesses were performed via high-grade commercial public-key end-to-end encryption. No other external operations were possible with the enclosure/computer other than the unplugging of its power cord. All administration services, login capabilities, web services, e-mail, etc. had been disabled and/or removed. Specifically, the computer was configured to execute one program and one program only. That is, the computer executed the single identity pattern-matching algorithm that had been review, tested, inspected, and unanimously approved by the policy/oversight body. For this pilot, this body consisted of a representative from each of the participating public health organizations aided by their respective IT staffs

As the reliability of this device and its correctness was of highest concern, the security configuration process was intentionally meticulous and comprehensive requiring the physical presence of an individual from each jurisdiction in order to make changes. Failure to accurately compute results or properly protect the information contained within would have rendered the device useless or even harmful, with significant loss of confidence from each of the participating organizations and their constituency. Operation of this privacy device prototype was intentionally very simple. In order for organizations to identify potential duplication issues, each generated a data file that contained a set of records for the individuals represented in their respective databases. For the initial pilot system, key fields included:

- Last name
- First name
- Date of birth
- Gender
- Ethnicity
- Social Security Number
- Local jurisdiction identifier

Using an agreed upon data file format for these records, each organization securely transferred its file to its respective directory on the privacy device computer. These directories were accessed in a "sally port" like fashion. That is, the organizations placed their data files within these directories, and then upon completion the Black Box algorithm removed the data files from these directories, decrypted the contents and transferred the resulting data into the local memory of the computer. In this manner, there was never any direct communication path between the algorithm and the external environment, as the existence of such paths would have provided a prime target for exploitation by an adversary.

Within the computer, the single program that was run continuously scanned each of the directories for new data files. When a new data file was detected, the file was carefully ingested and in-memory representation of the data is created. The source data file was then immediately securely deleted using multiple file re-writes. The directory scan and file ingest times were specifically engineered so that the sources data files, despite being encrypted, resided on the internal computer disk for a minimal amount of time (e.g. seconds).

In the event a new data file was received from an organization, the old representation was immediately discarded (erased from memory), and the new representation was then compared against the representations held for each of the other organizations. After the comparisons where completed, a report output file was prepared for each organization, identifying only those matches that are made with records of another organization. As they contained no private information, match files remained in the device directories until deleted by the respective organization (or whenever the privacy device system was restarted via power cycling). To further prevent PII exposure, the match files contained only the local organization's unique identifiers and no private source data fields. After a computation cycle, a participating organization was then able to use these identifiers to discuss possible lost-to-care or duplicate-care issues with the other corresponding organizations.

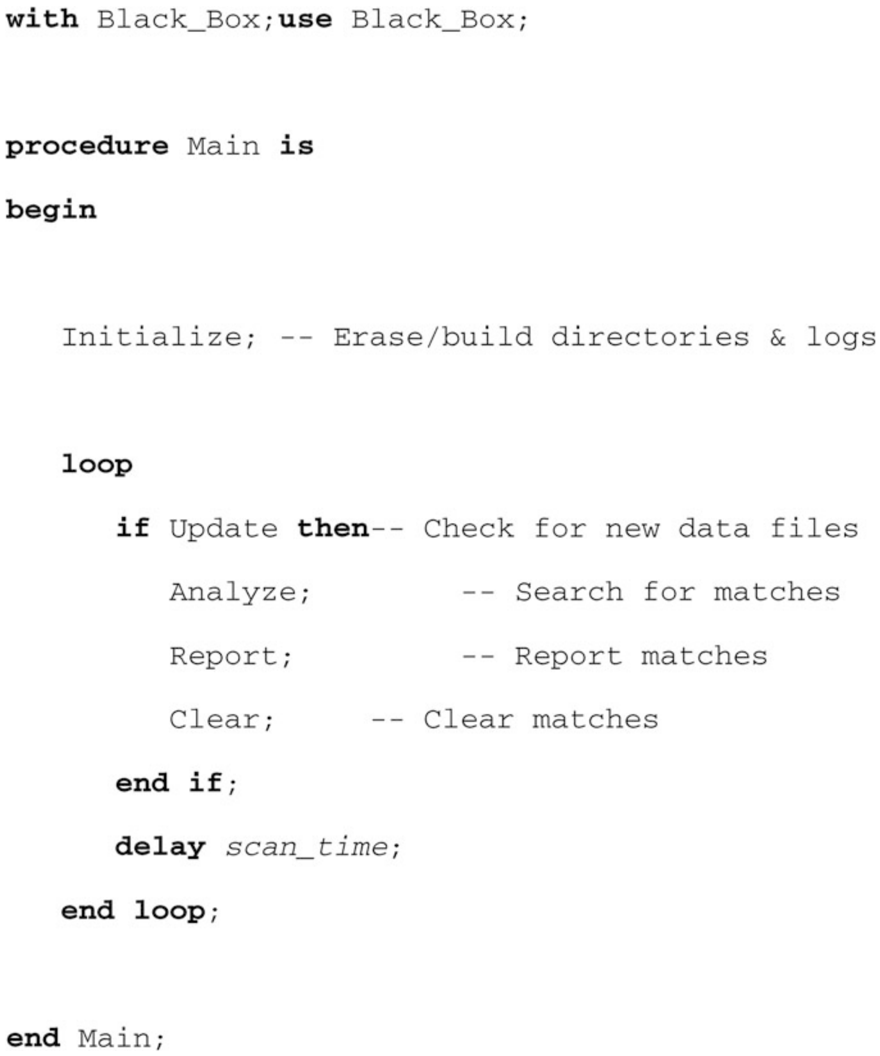
As the Black Box computer intentionally had neither a console nor display and was itself locked in cabinet without any remote monitoring capabilities, ascertaining the operating status of the device could only be performed by the participating jurisdictions. This was possible via a set of log files that was maintained by the algorithm for each jurisdiction. These logs contained the dates and times of ingested data files, when the matching process was performed, and summaries of the degree of matching found. Any errors detected in the input data file formats were reported back to the respective organization through this mechanism. Although operating within a Georgetown University computing facility, no member of the university staff had any ability to examine or monitor the status or contents of the device while it was in operation.

6.2 Pilot System Algorithm

Of all Black Box components, the item perhaps of greatest concern was arguably the algorithm contained within. From a reliability perspective, if this program were to have failed during the pilot's operation (e.g. as a result of an undetected programming error), the jurisdictions (or the developer) would not have had any way of knowing the cause. Although all data transmitted and stored was encrypted, such a failure could have conceivably resulted in a file containing PII persisting far beyond its expected (very short) lifetime upon the device. Such failures, however, could have also severely jeopardized each organization's confidence and trust in the device. If the device was not reliable, organizations would have been justifiably skeptical of its accuracy and its ability to protect such important information. The resulting loss of trust would have rendered the privacy device of little or no value, with the possibility of introducing harm via improper disclosure or wasted time pursuing inaccurate results. Thus, the reliability of the algorithm was of utmost importance throughout the process.

Providing added mechanisms for local real-time status display and remote diagnosis, however, would have increased the complexity of the design and the accompanying risk of compromise, exposing additional penetration paths that an adversary could have potentially exploited. During the development phase of this effort, a system complexity versus system integrity tradeoff became immediately prevalent in the discussion. Adding new features to the design to improve utility or operational use increased overall system complexity. With this added complexity came a tension upon the system's integrity. That is, the consideration of each new feature challenged the assurance of the system's impenetrability level. The pilot activity revealed that this complexity/integrity tradeoff is a fundamental, pervasive issue that must be recognized, addressed, and balanced throughout all phases of any Black Box system's lifecycle. For this pilot, the designers opted to maintain the highest level of simplicity whenever possible to aid the assurance process.

In accordance with its high-reliability and high-integrity design philosophy, the Ada programming language was selected for the algorithm specification and implementation (ISO/IEC 2012). Its unambiguous semantics, extremely strong type and constraint checking, exception protections, formally validated compilers, and overall reliability philosophy were key ingredients leading to this decision. The following is the main subprogram of the pilot system's algorithm:



```
with Black_Box;use Black_Box;


procedure Main is

begin


   Initialize; -- Erase/build directories & logs


   loop

      if Update then-- Check for new data files

         Analyze;        -- Search for matches

         Report;         -- Report matches

         Clear;     -- Clear matches

      end if;

      delay scan_time;

   end loop;


end Main;
```

As can been seen from above, the algorithm was kept very simple and consisted of a single infinite loop. The subprogram `Initialize` was used to create each organization's directory and corresponding log file should, they not already exist. If the directory did exist, its contents were erased, ensuring a fresh start. The package `Black_Box` contained the data structures that represented each organization's data set and the resulting cross organizational matches, along with the algorithm's operations that act upon them (`Update`, `Analyze`, `Report`, and `Clear`). Each of these subprograms was coded so that they would successfully complete, regardless of any internal error or exceptions that might result.

Of all the subprograms, `Update` was perhaps the most worrisome and complex as it involved the ingestion of external data files. While all organizations agreed to a single input format, the algorithm could make no assumptions regarding the input file's compliance as mistakes or errors could otherwise have rendered the system painfully inoperative. Thus when a new data file was detected within the `Update` subprogram, the new input file was very carefully parsed to ensure proper range values and format across all fields. In participating organizations' actual daily practice, it was not uncommon for their source databases to contain blank fields or legacy field formats that contained various wild card characters and special values for missing data elements (e.g. a birth year, but no birth month or day, or "000-00-0000" when a SSN is unknown). The `Update` subprogram's job was to reliably parse through all these various possibilities, reporting format errors back to an organization through its log file, ultimately creating a vector of properly type constrained person records for the corresponding organization. If the process was successful, `Update` returned a **`true`** value, allowing the algorithm to proceed. However, if an unrecoverable problem was detected, **`false`** was returned, preventing the subsequent matching and reporting operations from executing until a new data file was successfully received and processed from the organization.

With a successful (**`true`**) completion of `Update` subprogram, the remaining operations `Analyze` and `Report` were far less perilous as all data structures were now properly type checked and range constrained in comfortable mathematical fashion. The primarily role of the `Analyze` subprogram was to create a vector of records with persons that matched across all represented organizations. Match records contained values that identified the organization, their corresponding person unique identifiers, and a set of values that characterized which and how their fields matched including a score that indicated the likelihood that two individuals were actually the same. Scoring criteria was established via unanimous consent by the participating organizations during the algorithm design process, and then encoded into the `Analyze` subprogram.

The subprogram `Report` had very a predictable role and behavior, predominately creating the matching report output files for each of the organizations within their respective directory. To ensure no memory leaks over time (a common programing flaw), the `Clear` subprogram was used to properly release the dynamic data structures used in the matching process, before the entire process was repeated after a short specified time delay.

6.3 Pilot System Testing and Verification

As a Type 3 device, verification of the prototype system was undertaken using conventional software testing methods, manual code inspection, and comprehensive output file examination commensurate with best software engineering industry practice. Facilitated by participating organizations, a corpus of synthetic test data was used to test the algorithm under many diverse situations. As anticipated, the majority of programming flaws identified in the early testing phase were in the input process dealing with the external data files. However, once data was ingested and represented within the algorithm's strongly typed framework, no errors that would result in catastrophic failure (i.e. program crash or private information exposure) were detected. This was in part a testament to the oversight and involvement of the policy group in specifying and approving the algorithm's behavior. Thorough testing, however, did uncover an obscure programming logic flaw in the matching process due to an incorrect assumption regarding initial variable conditions. While conventional testing methods appeared adequate for a program of such modest size ($\sim$1000 lines), this process illustrated the critical importance of having a complete formal specification of the algorithm and the use of mathematical assertions and automated program proof-of-correctness techniques necessary to obtain a Type 2 or higher assurance level.

6.4 Pilot System Summary

The Black Box pilot system described here was heralded as a success (Ocampo et al. 2016). In total, the device processed well over 150,000 private information records identifying thousands of previously unknown matches with very high assurance. In total, the computation consumed approximately 20 min, a strong contrast to an otherwise manual process that would have easily extended beyond two years. More importantly, the process was executed entirely without any private information ever being revealed. The pilot exposed and illustrated the diverse spectrum of issues that must be responsibly addressed across a Black Box system's entire lifecycle, from initial design and procurement, to decommissioning and disposal. In summary, the system illustrated that the Black Box technique to private information sharing and analysis is both credible and viable. Moreover, the system successfully challenged the pervading assumption that analysts must have direct access to private, personal information to help further advance national security and public safety objectives. It illustrated that the tension perceived between personal liberty and these objectives need not exist. Rather, it demonstrated that security and safety goals can be met while simultaneously protecting personal information, and that such information need only ever be exposed to select individuals when there exists a very clear legal authority and established need.

7 Privacy Assurance Technology—Type 2

The pilot system discussed provided an illustrative example of an effective Type 3 system design and implementation. Observations throughout its develop process and end-to-end lifecycle helped identify the strengths and limitations of such systems. The technological basis of Type 3 systems is best industry practice. Candidly, as a system is scaled with increasing numbers of individuals and growing data volumes of ever increasing sensitivity, current best industry practice is simply not adequate given the evolving sophistication and insidious nature of contemporary adversaries. Evidence of this assertion can be witnessed each week with yet another major system compromise announced in the news media.

Assessing the pilot's Type 3 design, there are two areas of technical privacy concern. The first involves the method used to transfer private files into the Black Box. Configured using a private data sally port, direct access between the external environment and the internal algorithm is prevented. However, exposing computer file system directories to the outside world presents a potential exploitation path for an adversary, despite whatever firewall, encryption, and user access restrictions that might be imposed. The amount of software involved in a contemporary operating system's file management software and network data transfer applications often comprises many tens of thousands of lines of code (or far greater). Unfortunately, unless this code is specified, designed, and implemented perfectly, an adversary can potentially exploit any weaknesses that may have been overlooked (e.g. buffer overflows, range constraints, undefined states, etc.). As software systems increase in size, catching such mistakes becomes increasingly difficult and expensive. Alas, software "bugs" are indeed commonly found in software systems developed today despite earnest claims of best industry practice.

The second area of concern involves the method for specifying the Black Box algorithm. In the pilot system, while the algorithm was developed outside of box and available for all policy body members review and inspection, it eventually had to be compiled and installed in the Black Box prior to its sealing. This too presents a set of potential exploitation paths, as well as a very real logistical nightmare as the number of participating organizations is increased. Ensuring that the specified algorithm is correct and that the code transferred, installed, and ultimately run on the Black Box involves a large number of technical steps that must be carefully monitored and verified throughout. Unfortunately, this is a very complex process involving many more software modules with potentially hundreds of thousands of lines of code (or even millions). Assurance that this entire ecosystem is without exploitable flaws is far beyond best software engineering practice for any application beyond modest size. Further compounding scalability is the number of parties that would need to be involved to monitor, inspect, and ultimately be present to supervise the loading each time a new algorithm revision is needed becomes very impractical. To address these areas, several modifications are made to the pilot configuration in order to achieve a Type 2 assurance level, as shown in Fig. 4.

At the core of Fig. 4 is what is labeled as a "Secure" computer. As stated previously, perfect security is very elusive. However, the computing industry has made considerable progress developing computers and their companion operating systems for applications where high assurance is vital (e.g. avionics, power systems, medical equipment, etc.). A common framework for specifying computer security and assurance requirements exists and has been widely adopted (ISO/IEC 15408). While there can be no claim that these systems are totally without flaws, the

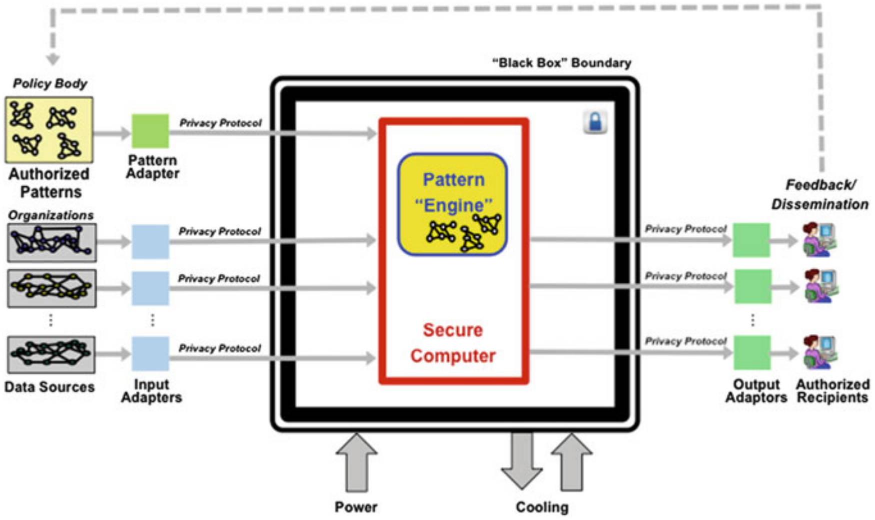


Fig. 4 Type 2 Black Box system design

development process is very rigorous, involving strict quality controls throughout that greatly boosts confidence. Of particular importance is the type and thoroughness of testing of every component and their interactions with others, yielding a certification and accreditation level with supporting documentation evidencing its rigor. While more costly than typical development efforts, this process is warranted by the policy body's assurance demands in order to responsibly mitigate an increased risk level.

Absent from Fig. 4 is the internal directory structure that the pilot system needed to expose to the outside environment. Transfer of bulk data (albeit encrypted) is a potential dangerous activity, as the contents of these data files could potentially contain malware that was injected by adversary. Thus, the data file transfer and file directory structure is replaced with a set of external input and output adapters that interface to the source data organizations and the result recipients, respectively. The primary role of these adapters is to convert source and result data into a set of transaction sequences that flow across the Black Box boundary. These transaction sequences are designed to move single data units, one at a time, verifying the format and validity of each. This is performed using a special privacy transfer protocol,[1] crafted specifically for high-assurance Black Box applications. This protocol is designed to enable all data transactions and related software handling components to be subject to mathematical proof-of-correctness rigor. This is possible with a

[1] The Hypergraph Transport Protocol (HGTP) under development at Georgetown University is specifically designed for this purpose.

complete protocol specification that is formally defined and verified with the inclusion of a vulnerability analysis that spans the full range of possible data values and transaction sequences.

At the right side of Fig. 4, pattern detection reports generated by the pattern engine flow out in a manner similar to the data input process, but in reserve order. That is, triggered pattern identifiers and the associated information identifiers exit the box via the privacy protocol. Once outside the box, this protocol is then converted to a form recognizable by an operator or alternatively to a form that can be processed by the contributing source organizations or investigating bodies that participate in the feedback/dissemination loop for oversight and compliance.

Lastly, rather than expose the internals of the Black Box to a new and potentially incorrect or vulnerable algorithm each time an analytic change is needed, a reusable pattern "engine" is used in Fig. 4 instead. This engine is itself a special algorithm, very carefully engineered to ensure that its pattern-matching operation cannot be modified in any fashion. It is coded one time, test, repaired, and verified perhaps multiple times, but then installed and authenticated in the Black Box once where it remains unchanged until the entire rigorous is repeated to accommodate new features. This process is critical for preventing any type of accidental or adversary-assisted disclosure of private information. Then henceforth, in place of transferring executable code to the Black Box, detection patterns expressed in a special analytic language[2] are instead transmitted, using the same privacy protocol for input data.

Inside the box, the engine interprets remotely specified pattern statements carefully versus trustingly executing them as in the Type 3 design. This interpretation step has the added security benefit that patterns expressed in the specification language cannot cause harm to the Black Box execution, given assurance in advance that the engine is correctly coded. With multiple participating organizations, the engine is configured so that the only patterns it will process are those that are properly expressed in the pattern language with all participating organizations simultaneously agreeing. Unanimous agreement is established by requiring each organization to send the specific pattern that they authorize to the Black Box where they are then compared against all the others. Internally, the engine only proceeds with data analysis and reporting when all of its received patterns are verified and are in proper agreement. Once developed, proven, loaded, and authenticated, the pattern engine algorithm within the Black Box cannot be modified without repeating the entire rigorous, monitored process. However, operational changes to how the Black Box behaves can be accommodated via updates to the pattern specification, considerably reducing the burden associated with refreshing the Black Box's internals.

[2] The ATra language under development at Georgetown University is specifically designed for this purpose.

8 Privacy Assurance Technology—Type 1

High-risk sharing and analysis applications involving extremely private personal information with large volumes of data about large number of individuals will invariably demand the highest level of privacy assurance—Type 1. This would likely include national or international applications that require the greatest level of protections in compliance with law and international treaty. To meet these highest assurances, several additional refinements are needed, as shown in Fig. 5.

At the core of Fig. 5 is now a "trusted" platform. In contrast to the Type 2 secure computer, this platform is a hardware/software device that has been designed and implemented in its entirety with thorough mathematical rigor to ensure its complete proof-of-correctness. As envisioned, this device would be a custom or specially tailored computing system specifically designed for this application. That is, features commonly found in typical off-the-shelf general-purpose computing systems that are not expressly needed to operate the pattern engine would be permanently disabled or removed from the design. Examples of superfluous items might include file storage machinery, all network channels, all input/output channels (excluding only that needed to support the privacy protocol), all display interfaces, and perhaps a large bulk of what is often resident in a typical operating system. In this scenario, the embedded system platform is designed, implemented, and verified precisely for this one privacy application at the greatest level of simplicity to ensure minimal exposure of vulnerability paths.

As software components beyond a few thousands lines of code are typically very hard to prove correct, all direct protocol communication with the Black Box is

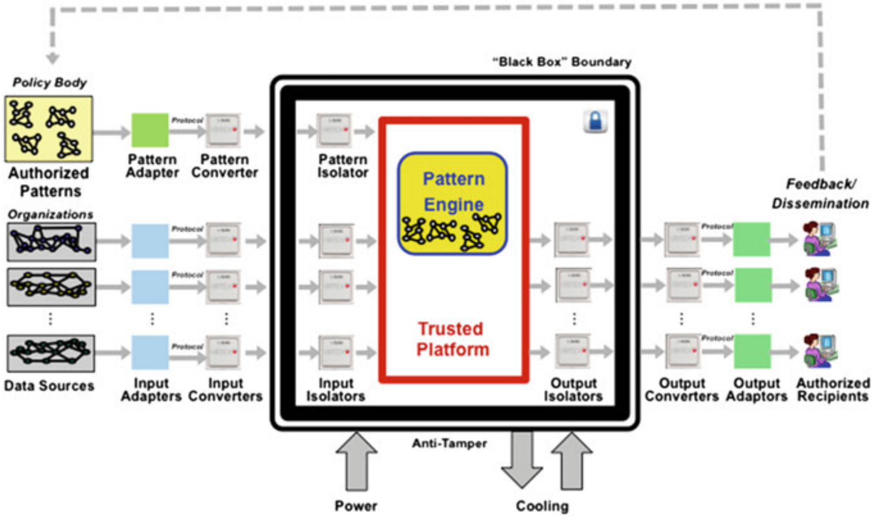


Fig. 5 Type 1 Black Box system design

replaced with a communication channel implemented directly in hardware using combinatorial logic components. This can be achieved with contemporary tradecraft leveraging technology items such as Field Programmable Gate Arrays (FPGA) and Application Specific Integrated Circuits (ASIC). These components have the advantage that their programming can be accomplished using a specification that is more readily subject to mathematical verification. In Fig. 5, each of the source organizations relays their private data to externally located input adapters. These adapters convert the data into a privacy protocol sequence. Each protocol sequence is then sent to an input converter that transforms the transaction into a discrete electrical or optical signal that crosses the Black Box boundary. Within the Black Box, this signal is fed directly to an input isolator that converts the signal into the digital transaction form needed by the pattern engine for processing. The same technique is applied for pattern specification, and the opposite process is used for outputting results to authorized recipients. The proper design and handling of isolators in this manner significantly reduces the vulnerability paths into and out of the Black Box, commensurate with the highest privacy assurance level.

Finally, to further strengthen the actual enclosure that encases all internal Black Box computing components, the use of "anti-tamper" techniques need be employed. This is an area of unique technical tradecraft that prevents adversaries from gaining physical access including special seals and alarm sensors. For fail-safe privacy operation, the primary purpose of these items is the removal of power from the internal components to ensure permanent, irretrievable loss of all Black Box data contents.

9 Operational Considerations

The privacy assurance approach advocated here was derived assuming existing, understood analytic tradecraft and proven, off-the-shelf technology components. While the myriad of technical issues is plentiful with the full spectrum of design and implementation aspects well beyond the scope of this publication, none of the constituent techniques and components described here are particularly new, distinctly novel, or technically unfounded. It is the careful configuration of these components and their unique operationalization within a privacy policy framework that is novel. The proper application, integration, and deployment of these techniques does require both a skilled workforce and a privacy work ethic that is often unfamiliar in everyday computing industry. The pilot activity discussed above helped reveal many of these characteristics. Beyond the technological realm, the primary process issues that must be addressed in tandem include:

- The establishment of a policy body and its associated processes for defining and authorizing patterns that would constitute reasonable suspicion/concern.

- The establishment of an oversight function or oversight body to monitor the operation of a Black Box configuration, including the auditing of input patterns and output reporting to ensure legal compliance.
- The establishment of specific development and deployment process procedures including design, implementation, configuration, physical and cyber protections, testing, and certifications of the Black Box *and its interfaces* to ensure sustained operational system integrity.
- The establishment of operational polices and procedures for identifying, protecting, and mitigating specific vulnerabilities across an end-to-end system deployment.

10 Conclusion

The material in this chapter outlines an approach that enables organizations to share and analyze information in a manner that respects and embraces individual privacy rights. Although discussed here within a privacy policy context, the Black Box assurance approach is applicable to a diverse spectrum of information sharing and analysis challenges including:

- Commercial or community organizations desiring to protect sensitive information across their organizational boundaries to enable cost savings and operating efficiencies while providing improved services.
- Compartmented organizations needing to protect classified or highly sensitive information on a strict need-to-know basis yet work collaboratively towards shared objectives.
- International organizations needing to protect highly sensitive information, perhaps of a treaty, compliance, or deterrence nature, yet work cooperatively to identify areas with common goals and interests.
- Numerous other applications, ranging from health records management and HIPAA compliance to financial information processing for waste, fraud and abuse detection.

With the acceptance and adoption of the Black Box approach to privacy assurance, a new paradigm for private information sharing and analysis is poised to emerge. With this technique, it is possible to declare that all private information about an individual must remain hidden from *all* other individuals unless there is explicit permission for disclosure from the information's owner (e.g. application for a car loan), or legal authority for its examination (e.g. criminal investigation). Organizations that curate private information (e.g. a credit card company) would do so by storing this information in a Black Box container, hidden from view from any of its members. In Black Box fashion, the organization could perform considerable analysis upon this data using the pattern specification and policy body approval machinery without the private data entering an employee's view, except perhaps at initial entry. For many applications, reference to an individual can be accomplished with the organization's assigned identifier, versus repeated exposure of name, address, social security number, telephone number, etc. Computationally, this has the added potential benefit for reducing ambiguity errors where individuals are confused due to similar personal information (e.g. their names).

This construct suggest a further generalization where a set of black boxes may be configured to interact with each other to further reduce the visibility of private information. While this does not suggest that removing human inspection and approval from decision processes is always possible or appropriate, it does offer a mechanism that can greatly reduce private information exposure, limiting access to only those individuals based on confirmed, authoritative need. This new paradigm enables a richness of private information to be shared and analyzed when needed, yet allows carefully restricted user access based on a need-to-know, authorized-to-know basis, and allowed-to-know basis.

Further exploration of these techniques and their many variants offers a unique hope towards addressing the otherwise difficult tension between Security and Liberty. A successful resolution of this tension has profound implications on the modern Information Age.

References

Campos, J. (2014). Civil liberties and national security: The ultimate cybersecurity debate. *Homeland Security Today*, January 27, 2014.

Center for Strategic and International Studies. (2014). *Balancing security and civil liberties—Principles for rebuilding trust in intelligence activities*, Washington D.C., May 15, 2014.

Committee on National Security Systems. (2010). *National information assurance (IA) glossary*. CNSS Instruction No. 4009, April 26, 2010.

Gilmore, J. (2014). Balancing homeland security and civil liberties. *Washington Times*, March 6, 2014.

ISO/IEC 15408. The common criteria for information technology security evaluation

ISO/IEC 8652. (2012). Ada reference manual—Language and standard libraries, 2012(E).

Smart, J. C. (2011). Privacy assurance, international engagement on cyber. Georgetown Journal of International Affairs. http://avesterra.georgetown.edu/tech/privacy_assurance.pdf

Ocampo, J. M., Smart, J. C., et al. (2016). Developing a novel multi-organizational data-sharing method to improve HIV surveillance data for public health action in metropolitan Washington DC. Journal of Medical Internet Research Public Health and Surveillance.