MIT COMPUTER SCIENCE AND ARTIFICIAL INTELLIGENCE LABORATORY

Needles in Haystacks: Creating Information Balance Sheets for Personal Data

Testimony of
Daniel J. Weitzner <djweitzner@csail.mit.edu>
Director, MIT Decentralized Information Group
Principal Research Scientist, MIT Computer Science and Artificial Intelligence Laboratory
http://dig.csail.mit.edu/2013/07/pclob-weitzner-accountability.pdf
Testimony as originally presented:
http://dig.csail.mit.edu/2013/07/pclob-weitzner-accountability-orig.pdf

Before the United States Privacy and Civil Liberties Oversight Board
Workshop Regarding Surveillance Programs Operated Pursuant to Section 215 of the USA PATRIOT Act and Section 702 of Foreign Intelligence Surveillance Act
July 9, 2013

I. Introduction

Traditional legal and technical approaches to privacy protection show their limitations, as pervasive collection, digital storage and analysis of personal data goes from the exception to the norm. New information privacy challenges in both the private and government sectors arise from the fact that. In the past, extra effort was required to collect sensitive information leading to a natural bias toward privacy. As the interaction between the government and private sector organizations with respect to both telephone metadata (the 215 programs) and Internet content and metadata (the 702 programs) illustrate, government requests for very large amounts of personal data – such as all telephone metadata generated by a single network operator – are easy to satisfy. However, the technical challenges associated with reliable and trustworthy oversight of these programs are not yet well addressed.

This testimony reviews the legal and technical challenges of establishing accountable information usage, discusses current technical developments in the field, and offers the following recommendations:

1. Establish clear rules for how personal information can be used.
2. Require publicly visible *Information Balance Sheets* that report on how personal data is used in law enforcement and national security investigations.
3. Use automated policy analytics to assist courts in their independent oversight function.

Technical advances in computer science and artificial intelligence have increased our analytic capability to detect threats and solve crimes by combing through large volumes of personal data. This data can be thought of as the haystack, inside of which may be hiding a needle: a single piece of data which could be the clue to stopping a terrorist act about to happen or the evidence necessary to convict a criminal of a crime. At the same time, the volume of personal data collected and the complexity of analytics applied to those data poses new challenges for the institutions of government responsible for assuring accountability to rules designed to protect our civil liberties. In other words, how can we monitor the process of sifting through the proverbial haystack?

We no longer expect law enforcement investigators or national security analysts to run their investigations with hand written notes on index cards. Instead, we provide increasingly sophisticated automated investigative analytics to help find the needle in that haystack. By the same token, if we are to assess accountability to rules governing use of personal information, we need equivalently robust computational power to monitor these systems. We need systems that can answer the question whether government agencies are adhering to the strict contours of the law or making use of personal data beyond what is authorized. In other words, we want to be able to tell whether an agency is using a magnet to extract the needle and nothing else, or a pitchfork, pulling along with it a lot of hay. Recent advances in computer science research on accountable systems show that it is possible to verify compliance with privacy rules using computational techniques that can operate at large scale.

At their best, well-designed information systems contribute transparency and clarity to those who rely on them. Over the last five years, many around the world have recognized the ways in which online information can open up government and private sector institutions with transparency tools. We should bring that same spirit to work in the realm of privacy protection. Much work needs to be done to deploy these systems, but they are the only means by which we can both allow intelligence agencies to conduct aggressive hunts for needles and at the same time offer meaningful transparency to assure the public that those needles are being extracted in a manner that respects our basic civil liberties.

II. Accountability Requirements in Surveillance Programs with Broad Collection Authority

A. The 'Big Data' Privacy Challenge

Here is the central accountability challenge posed by large-scale surveillance programs: agencies of the government are entrusted with possession of large amounts of personal data on the promise that will only use it in a legally permissible manner. As DNI General Counsel Robert Litt recently explained:

> "In 2012 fewer than 300 identifiers were approved for searching this [telephone metadata] data. Nevertheless, *we collect all the data because if you want to find a needle in the haystack, you need to have the haystack*, especially in the case of a

> terrorism-related emergency, which is – and remember that this database is only used for terrorism-related purposes."[1]

Recognizing that there is considerable debate about whether the "relevance" standard in Section 215 of the Patriot Act properly justified access to wholesale datasets such as *all* telephone metadata from a particular network, we should also acknowledge that the intelligence community has authority and the legitimate need to collect very large volumes of personal data, even if not all data. Therefore, the core legal, technical and administrative question is whether there is adequate oversight of the subsequent *use* of that data.

In the public debate that has ensued since the scale of scope of these programs has become better known, some argue[2] that we need new substantive rules to limit the conditions under which government can access or use such personal data. Others suggest that the legal rules are adequate but that a greater degree of transparency and accountability is needed to guard against abuse and assure the public that the rules are actually being followed[3]. Hardly anyone has suggested both that the rules are adequate and that we have sufficiently accountable oversight mechanisms in place.

B. Special accountability mechanisms required for assessing compliance with *ex post facto* usage rules

Rules put in place by Congress and the FISA court govern the use of personal data *after* it has been obtained by the government. In defending access to telephone and email metadata, officials point out that the relevant legal authorities prohibit analysts from actually querying data on US persons without proper predication and a court order. Furthermore, in most cases the data can only be used for terrorism investigations. In the last month we have heard much discussion of internal controls put in place to assure compliance with statutory rules, FISC orders and internal policies. Those mechanisms are no doubt important, but are not sufficient to provide adequate transparency for rules that govern information usage.

Monitoring data usage is far more complex as a technical matter than monitoring access or collection. Internal audit mechanisms must be able to reliably report on how data is used within an institution after the initial collection event. Various techniques such as access logs and segregated databases have been suggested or put in place to meet transparency and accountability needs. While valuable, they do not offer sufficient information to demonstrate compliance with usage rules. First, access logging – the ability to record which individual analyst has actually requested access to a particular piece of data – can only

[1] Remarks at Newseum, Special Program - NSA Surveillance Leaks: Facts and Fiction Wednesday, June 26, 2013. (emphasis added)

[2] "Groups to sue over NSA surveillance," USA Today, July 8, 2013

[3] "It is up to Congress, the courts and the public to ask the tough questions and press even experienced intelligence officials to back their assertions up with actual evidence, rather than simply deferring to these officials' conclusions without challenging them." Wyden/Udall statements on disclosure of bulk email records collection program. (July 2, 2013)

track who accesses a piece of data, not what that individual actually does with the data. Logging and auditing access is an important component of any internal security system and may reveal circumstances in which an individual user is improperly viewing a piece of data. Still, such logging will not reveal violation of usage rules. Second, data obtained through surveillance orders may be stored in segregated databases. Such controls may help discourage analysts from improperly combining data, but these approaches only segregate the data, not the individual analysts and therefore do not provide any check on possible onward use of that data.

C. Public Information Balance Sheets for Personal Data – How to audit classified activities to produce public trust.

Systems designed to produce accountability for data usage rules in a national security context face the unique challenge of having to respect the security classification of much of the data, while at the same time generating suitable independent and publicly-trustable audit trails. Needless to say, we cannot expect intelligence agencies to declassify data in any reasonable timeframe to demonstrate that that it is used consistent with the laws. At the same time, operating surveillance programs collecting data of ordinary citizens not themselves subject of any particularized suspicion, we ought to require some evidence that this data is used in strict compliance with rules. The current approach to accountability for classified activities keeps the entire chain of data usage – from judicial authorization, to internal controls and audit logs – entirely classified, away from public scrutiny. There are accountability models that strike a more transparent balance between secrecy and oversight without compromising sensitive information.

Financial accounting standards offer an example of how information systems can give the public confidence in the behavior of institutions bound by specific rules without having to disclose proprietary information. The public, the markets, and regulators generally trust financial statements such as balance sheets and profit and loss tables because they are prepared according to a known set of rules that, if followed, produce consistent and reliable results. The integrity of this system depends not just on clear rules, but also on regular audits by trusted and independent professionals. Of course, inaccuracy can emerge due to either mistake or fraud. But on the whole, the financial accounting system has produced an enviable level of trust and confidence in a fast-moving, highly decentralized market system, in which each participating institution places a very high value on preserving the secrecy of core operating data. Advances in computer science research in the field of accountable systems suggest that it is possible to achieve a similar degree of confidence and secrecy in the operation of large systems analyzing personal data.

III. Accountable Systems Architecture to Measure Compliance with Usage Rules

Can systems that analyze large volumes of personal data also be designed to analyze whether the data in the systems is beginning used according to the applicable laws and policies? A growing community of computer science researchers has been working on the design of what we call accountable systems – information systems that are able to represent legal rules in computational format and then apply those rules to audit or transaction logs that record how data is used in those systems. Accountability is generally defined by computer scientists as the ability to hold an entity, such as a person or organization, responsible for its actions[4] or the ability to punish someone when rules are violated.[5] Those working in the field have shown how to apply these techniques to healthcare[6], law enforcement information sharing[7], copyright law,[8] and general designs that would augment the basic architecture of the World Wide Web to provide for more accountable information flow.[9]

A. Accountable Systems In Action

Research on accountable systems architectures in my lab at MIT has demonstrated that is possible to build systems that provide 'information accountability'[10] – the ability to pinpoint improper use of information as defined by legal rules expressed in machine-readable format. Figure 1 shows a system we built modeling a Massachusetts law prohibiting denial of public services based on individual health status. Our prototype analyzes a log of information used in this particular system and assesses those uses against a set of rules expressed in a specialized rule language. Expressing legal rules in this language enables us to use it somewhat like a programming language, allowing computation on audit log data to test policy compliance. We model a scenario in which a customer service representative for a hypothetical local telephone company is in possession of information suggesting that a customer may have a communicable disease. Seeking to protect phone company workers, the service representative denies a request by the customer to have a repair person fix the customer's home phone. This is an example of a policy whose restrictions are based on usage rules, not access or collection rules. The phone company is in legitimate possession of information about the customer's health status but is nevertheless not allowed to use it for determining service eligibility.

[4] Lampson, B. (2005, October). Accountability and freedom. In Cambridge Computer Seminar, Cambridge, UK.

[5] Feigenbaum, J., Hendler, J. A., Jaggard, A. D., Weitzner, D. J., & Wright, R. N. (2011, June). Accountability and deterrence in online life. In Proceedings of the 3rd International Conference on Web Science, ACM.

[6] DeYoung, H., Garg, D., Jia, L., Kaynar, D., & Datta, A. (2010, October). Experiences in the logical specification of the HIPAA and GLBA privacy laws. In *Proceedings of the 9th annual ACM workshop on Privacy in the electronic society* (pp. 73-82). ACM. And Lam, P. E., Mitchell, J. C., & Sundaram, S. (2009). A formalization of HIPAA for a medical messaging system. In *Trust, Privacy and Security in Digital Business* (pp. 73-85). Springer Berlin Heidelberg.

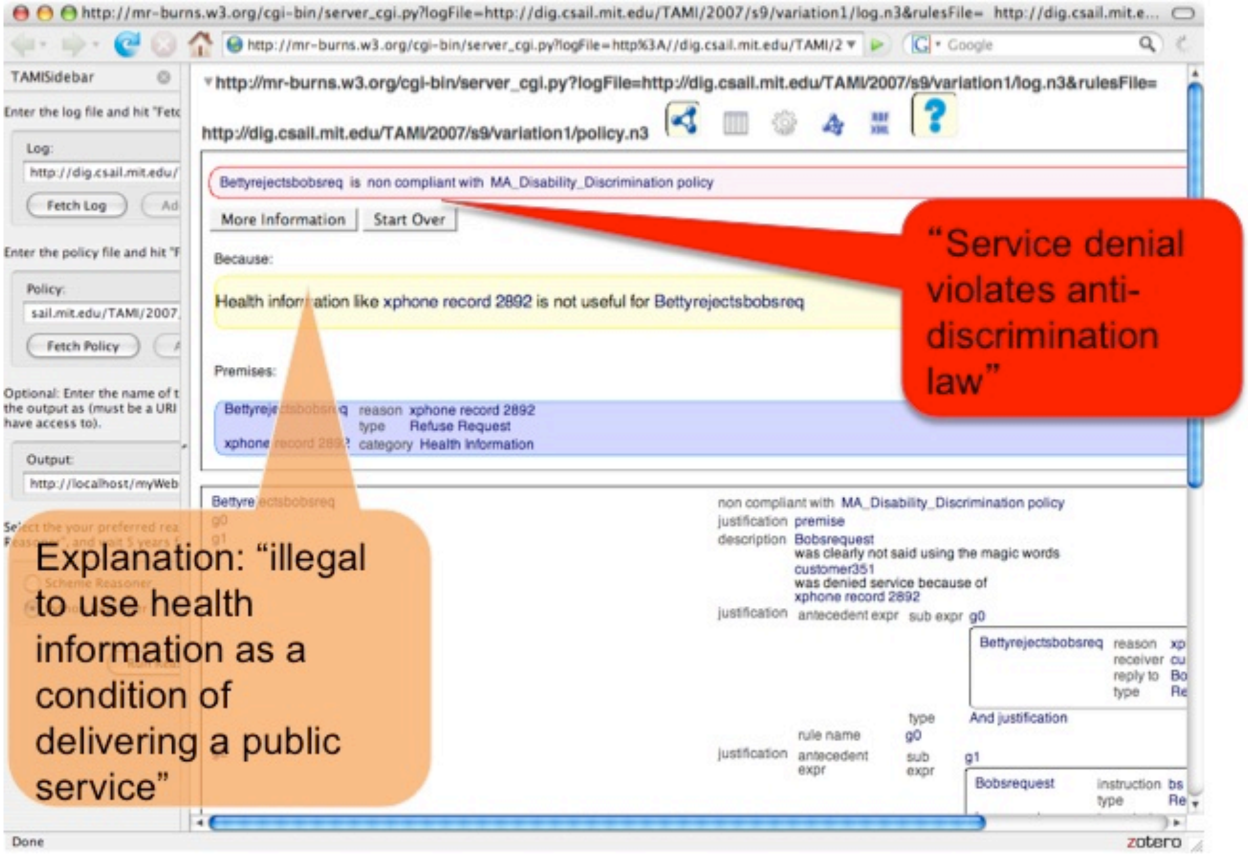
[7] Waterman, K. K., & Wang, S. (2010, November). Prototyping fusion center information sharing; implementing policy reasoning over cross-jurisdictional data transactions occurring in a decentralized environment. In *Technologies for Homeland Security (HST), 2010 IEEE International Conference on* (pp. 63-69). IEEE.

[8] Seneviratne, O., Kagal, L., Weitzner, D., Abelson, H., Berners-Lee, T., & Shadbolt, N. (2009). Detecting creative commons license violations on images on the World Wide Web. *WWW2009, April*.

[9] Seneviratne, O., & Kagal, L. (2011). Usage Restriction Management for Accountable Data Transfer on the Web. In *IEEE International Symposium on Policies for Distributed Systems and Networks (IEEE Policy 2011)*.

[10] Weitzner, D. J., Abelson, H., Berners-Lee, T., Feigenbaum, J., Hendler, J., & Sussman, G. J. (2008). Information accountability. *Communications of the ACM*, *51*(6), 82-87.

The legal rules models in this scenario are not applicable, of course, to the intelligence agency activity under discussion today. Still, our system demonstrates the ability to express and audit against rules governing the *use* of personal information. This is in contrast to features commonly found in systems that control and perhaps even create audit logs of *access* to data. To the extent that privacy rules governing intelligence activities have a similar structure, seeking to control the ultimate use of data, these systems described constitute an proof-of-concept of an approach to accountability to usage rules generally.



http://mr-burns.w3.org/cgi-bin/server_cgi.py?logFile=http://dig.csail.mit.edu/TAMI/2007/s9/variation1/log.n3&rulesFile= http://dig.csail.mit.edu/TAMI/2007/s9/variation1/policy.n3

Bettyrejectsbobsreq is non compliant with MA_Disability_Discrimination policy

Because:

Health information like xphone record 2892 is not useful for Bettyrejectsbobsreq

Premises:

Bettyrejectsbobsreq reason xphone record 2892
type Refuse Request
xphone record 2892 category Health Information

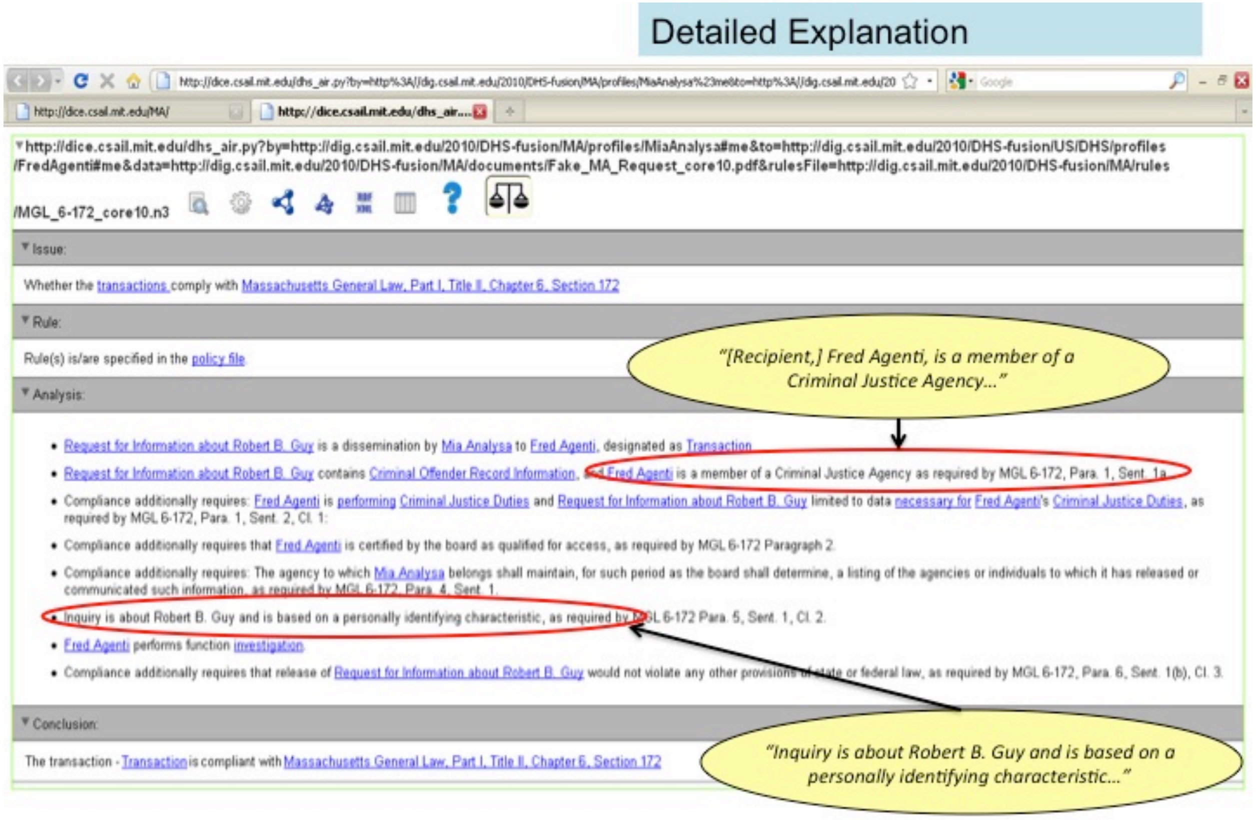
non compliant with MA_Disability_Discrimination policy
justification premise
description Bobsrequest was clearly not said using the magic words customer351 was denied service because of xphone record 2892
justification antecedent expr sub expr g0

TAMISidebar
Enter the log file and hit "Fetc
Log:
http://dig.csail.mit.edu/
Fetch Log
Enter the policy file and hit "F
Policy:
sail.mit.edu/TAMI/2007
Fetch Policy
Optional: Enter the name of t the output as (must be a URI have access to).
Output:
http://localhost/myWeb
More Information
Start Over
Google
Bettyrejectsbobsreq reason receiver reply to type
And justification
type g0
rule name
justification antecedent expr
sub expr g1
Bobsrequest instruction type
Done
zotero
"Service denial violates anti-discrimination law"
Explanation: "illegal to use health information as a condition of delivering a public service"

Figure 1 - Detecting violations of Mass. Anti-Discrimination Law

The red balloon highlights the policy analysis conclusion reached by the system – that the decision to deny this particular customer service is a violation of the Commonwealth's anti-discrimination law. Our systems are also able to provide an explanation of the legal conclusion reached. In this case, the orange balloon shows that the service denial is illegal because the law prohibits the use of health information as a basis for providing public services such as telephone service. The ability to offer an explanation for policy conclusions can be helpful as a just-in-time warnings for users to be aware when the action they are about to take might violate the rules in the system. Of course, if they continue with the action, the misuse could be logged in the systems audit system.

We have applied similar accountable systems technology to a prototype designed to help analysts in law enforcement-intelligence fusion centers to assess when they are allowed to share information with another agency in the fusion center. Figure 2 shows the accountability mechanism operating with a provision of Massachusetts criminal law that controls when investigative information may be shared with others. Here the act of sharing a piece of data is found to be compliant with the relevant law because the proposed recipient meets the statutory definition of a criminal law enforcement agency and the request is limited to a specifically identified individual per the requirements of the law. In this case the system analyzes the proposed action against the relevant legal rules and returns an answer with an explanation highlighting those items in the transaction log that a determinative in the policy reasoning.



Detailed Explanation

http://dice.csail.mit.edu/dhs_air.py?by=http%3A//dig.csail.mit.edu/2010/DHS-fusion/MA/profiles/MiaAnalysa%23me&to=http%3A//dig.csail.mit.edu/20
Google
http://dice.csail.mit.edu/MA/
http://dice.csail.mit.edu/dhs_air....

http://dice.csail.mit.edu/dhs_air.py?by=http://dig.csail.mit.edu/2010/DHS-fusion/MA/profiles/MiaAnalysa#me&to=http://dig.csail.mit.edu/2010/DHS-fusion/US/DHS/profiles/FredAgenti#me&data=http://dig.csail.mit.edu/2010/DHS-fusion/MA/documents/Fake_MA_Request_core10.pdf&rulesFile=http://dig.csail.mit.edu/2010/DHS-fusion/MA/rules/MGL_6-172_core10.n3

Issue:

Whether the transactions comply with Massachusetts General Law, Part I, Title II, Chapter 6, Section 172

Rule:

Rule(s) is/are specified in the policy file.

Analysis:

- Request for Information about Robert B. Guy is a dissemination by Mia Analysa to Fred Agenti, designated as Transaction
- Request for Information about Robert B. Guy contains Criminal Offender Record Information, and Fred Agenti is a member of a Criminal Justice Agency as required by MGL 6-172, Para. 1, Sent. 1a
- Compliance additionally requires: Fred Agenti is performing Criminal Justice Duties and Request for Information about Robert B. Guy limited to data necessary for Fred Agenti's Criminal Justice Duties, as required by MGL 6-172, Para. 1, Sent. 2, Cl. 1:
- Compliance additionally requires that Fred Agenti is certified by the board as qualified for access, as required by MGL 6-172 Paragraph 2.
- Compliance additionally requires: The agency to which Mia Analysa belongs shall maintain, for such period as the board shall determine, a listing of the agencies or individuals to which it has released or communicated such information, as required by MGL 6-172, Para. 4, Sent. 1.
- Inquiry is about Robert B. Guy and is based on a personally identifying characteristic, as required by MGL 6-172 Para. 5, Sent. 1, Cl. 2.
- Fred Agenti performs function investigation.
- Compliance additionally requires that release of Request for Information about Robert B. Guy would not violate any other provisions of state or federal law, as required by MGL 6-172, Para. 6, Sent. 1(b), Cl. 3.

Conclusion:

The transaction - Transaction is compliant with Massachusetts General Law, Part I, Title II, Chapter 6, Section 172

"[Recipient,] Fred Agenti, is a member of a Criminal Justice Agency..."

"Inquiry is about Robert B. Guy and is based on a personally identifying characteristic..."

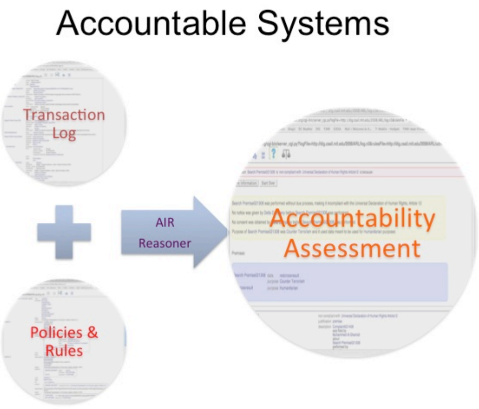
Figure 2 - Information Sharing Rules Compliance Guide

The user interface shown in Figure 2 presents an entirely computer-generated analysis of the policy compliance in a form familiar to lawyers, identifying the legal Issue being analyzed, the Rule being applied, an Analysis of the reasoning steps, and the legal Conclusion. We do not expect that this system will obsolete the need to teach law students the IRAC case briefing model. Rather, we have used this structure so that lawyers using this tool will find the information more accessible.

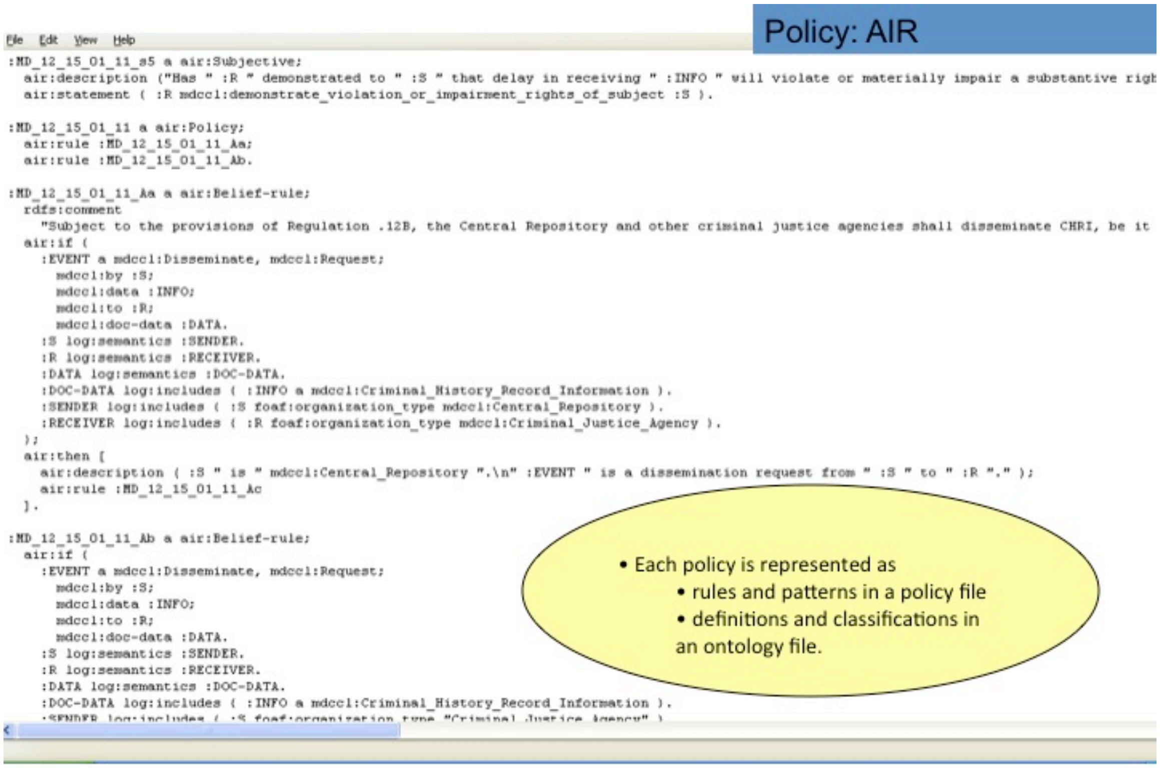
B. Accountable Systems Architecture

Each of the systems shown here are applications of the same general purpose infrastructure, consisting of three main components:

1. Policy language – a computer language specially designed to express legal rules in a form so that they can be applied to events in a transaction or audit log.
2. Reasoner – a system able to draw logical conclusions about how the particular legal rules expressed in the policy language apply to a set of transactions described in an audit log.
3. Justification user interface – a web-based interface that interprets the computation from the reasoned and provides an accountability assessment.



This basic set of system functions is designed so that it can be deployed in any system with regular logging of information usage. The policy language (see Figure 3 for a sample) is



```
:MD_12_15_01_11_s5 a air:Subjective;
  air:description ("Has " :R " demonstrated to " :S " that delay in receiving " :INFO " will violate or materially impair a substantive rigt
  air:statement ( :R mdccl:demonstrate_violation_or_impairment_rights_of_subject :S ).

:MD_12_15_01_11 a air:Policy;
  air:rule :MD_12_15_01_11_Aa;
  air:rule :MD_12_15_01_11_Ab.

:MD_12_15_01_11_Aa a air:Belief-rule;
  rdfs:comment
    "Subject to the provisions of Regulation .12B, the Central Repository and other criminal justice agencies shall disseminate CHRI, be it
  air:if (
    :EVENT a mdccl:Disseminate, mdccl:Request;
      mdccl:by :S;
      mdccl:data :INFO;
      mdccl:to :R;
      mdccl:doc-data :DATA.
    :S log:semantics :SENDER.
    :R log:semantics :RECEIVER.
    :DATA log:semantics :DOC-DATA.
    :DOC-DATA log:includes ( :INFO a mdccl:Criminal_History_Record_Information ).
    :SENDER log:includes ( :S foaf:organization_type mdccl:Central_Repository ).
    :RECEIVER log:includes ( :R foaf:organization_type mdccl:Criminal_Justice_Agency ).
  );
  air:then [
    air:description ( :S " is " mdccl:Central_Repository ".\n" :EVENT " is a dissemination request from " :S " to " :R "." );
    air:rule :MD_12_15_01_11_Ac
  ].

:MD_12_15_01_11_Ab a air:Belief-rule;
  air:if (
    :EVENT a mdccl:Disseminate, mdccl:Request;
      mdccl:by :S;
      mdccl:data :INFO;
      mdccl:to :R;
      mdccl:doc-data :DATA.
    :S log:semantics :SENDER.
    :R log:semantics :RECEIVER.
    :DATA log:semantics :DOC-DATA.
    :DOC-DATA log:includes ( :INFO a mdccl:Criminal_History_Record_Information ).
    :SENDER log:includes ( :S foaf:organization_type "Criminal Justice Agency" ).
```

- Each policy is represented as
  - rules and patterns in a policy file
  - definitions and classifications in an ontology file.

Figure 3 - Law expressed in AIR policy language

designed to express a wide variety of legal rules. Finally, our entire system is built with Semantic Web, linked data technology, a set of Web technical standards that enable the policies to be written in a manner that they can easily refer to a wide range of data types. Use of linked data techniques enables us to encode any given law or rule in the AIR policy language once and then apply that rule in a number of different systems, saving implementation time and ensuring consistent application of rules from one system to another.

C. Complementary Roles of Law and Technology

Just as new information technology raises questions about the substantive scope and enforceability of current law, different versions of the very same technologies can be put in service of more effective enforcement. This will increase public trust and institutional certainty. The accountable systems technologies described here can help institution's internal compliance efforts. And the capability to deploy computer-assisted accountability is critical to assessment of rule compliance when the scale of information transactions outpaces the ability of purely manual oversight. Indeed, there are a number of roles that machine-assisted accountability can play in enforcement, but none are a silver bullet that will magically decide what the proper scope of information collection and usage is for national security purposes. Those questions fall squarely in the hands of our courts and legislatures. Policy makers ought to be aware of these information accountability techniques and encourage their use. However, there should be no illusion that the mere existence of the technologies answer substantive policy questions raised by new surveillance and analytic power.

IV. Applying Accountable Systems Architecture to current surveillance programs

As the ease of data collection continues to grow, rules governing the usage of that personal data will be increasingly important to privacy protection. Of course, constitutional and legislative determinations will establish the upper bounds on how much data can be collected under different circumstances, but the size of the haystack is likely to be large and grow larger in the future. Usage rules feature prominently at the center of the current debate over 215 and 702 programs. Consider these two usage restrictions

1. Personal data from wholesale collection of telephone metadata will only be queried with specific predication.
2. Personal data from telephone metadata will only be used for terrorism investigations.

Adherence to both of these rules can make the difference between targeted selection of data with minimal intrusion on individuals for whom there is no articulable suspicion of wrongdoing, as opposed to a general search through data covering a large percentage of the population. Accountable systems with thorough logging of each information usage event and policy-driven analysis of that log data could both help on several fronts. First, real-time policy analysis of queries conducted by analysts can help warn individuals when they are engaged in what may be rule violations. Helping well-meaning data users to do the right thing ought to be a high priority. Second, data usage can be logged and analyzed for subsequent internal and independent oversight. Accountable systems reasoners can be used to analyze data from logs to detect possible rule violations. Finally, rigorous computational accountability techniques can be developed such that some part of the accountability assessment could be made public without exposing classified data. Careful design will be required here to avoid disclosing intelligence sources and methods, of course.

Experience from other accountability efforts, such as the financial realm, establish that these new accountable systems will not detect all rule violations. However, as with any other well-established auditing technique used today, computational accountability can provide a structured basis for scrutinizing activity in order to encourage the highest standards of institutional behavior and build public trust.

Our research results on accountable systems give us confidence that it is possible to deploy these techniques at large scale in operational environments. Basic and applied research by a number of research groups supported by the National Science Foundation, IARPA and the Department of Science and Technology Science and Technology Directorate have helped lay a strong technical foundation for these systems. However, to the best of our knowledge, these tools are not yet available for off-the-shelf deployment. Increasingly widespread use of access logs is a good first step on the path to widespread deployment of accountable systems, but as with most information technology, the marketplace will only respond with products and services to the extent that users, and those who oversee those users, indicate a need for the products.

V. Recommendations

As the PCLOB considers how to approach its own oversight of civil liberties, and as the Board formulates broader public policy frameworks for addressing these issues, we offer specific recommendations on how to take maximum advantage of the power of accountable systems technology to support civil liberties and increase public trust.

1. *Establish clear laws and policies setting concrete and objective rules regarding use of personal information.* Broad rules subject to a variety of interpretations risk inconsistent application, uncertainly on the part of data users, and loss of trust from the general public. For all of the power of information technology, there is no computer system capable of intuiting how vague rules should to be applied to specific situations. So, policymakers ought to work to establish the clearest possible parameters for information usage and be sure that those rules are expressed in a way that they violations of the rules are detectable.
2. *Require publicly visible Information Balance Sheets that report on how personal data is used in law enforcement and national security activities.* The new frontier of privacy protection will be enabled by systems that provide trustworthy, concrete evidence that the actual uses of information comply with all of the relevant rules.

Our work on accountable systems technology has shown that this is possible. We see this as a necessary addition to the Privacy Act paradigm of System of Records Notices and other privacy policy statements. These policy statements have played a vital role in institutional transparency, but are no longer sufficient, on their own, to support public trust. We should add to the current privacy model an expectation that all systems using personal data produce what we call Information Balance Sheets. Information Balance Sheets can attest to the nature of actual information flow and use in any given institution. Information Balance Sheets are similar to financial balance sheets that attest to the financial status of an organization. The public can rely on financial balance sheets because they are produced according to a reliable and provable consistent methodology such as generally accepted accounting practices. We can create a similar degree of certainty and trust with the handling of personal data by requiring all institutions handling personal data to produce Information Balance Sheets. Our research has shown that the same technology that produces Information Balance Sheets can also help individuals in an institution to do the right thing when using personal data. We also recommend that the PCLOB encourage national security and law enforcement agencies to build such policy analytic capabilities into the systems.

3. *Use automated policy analytics to assist courts in their independent oversight function.* Courts and other enforcement bodies charged with providing independent oversight of intelligence and law enforcement activities should investigate how to use accountable systems technologies to guide their enforcement efforts. The scale and scope of data analysis makes it unlikely that any oversight body can effectively assess rule compliance without such technical tools.

VI. Conclusion

As more and more of our public and private lives are recorded in digital information systems, the size of the haystack through with intelligence analysts will have to search will only grow larger. A central concern of the public and oversight bodies will be to assure that those who comb through these haystacks in search of needles are doing so with tools that act more like magnets than pitchforks. Magnets can extract the needle without also attracting the irrelevant hay. Those who set the legal rules governing these activities will have to be as precise as possible about what data can be collected and how it can be used. As a technical and operational matter, the ability to measure whether these rules are being followed will require computational tools that match the scale and sophistication of the underlying investigative systems. Information accountability techniques described here can bring to bear the analytic power of computer systems in a manner that provides basic transparency into the legal and policy implications of these complex investigative techniques for both independent overseers and the public, without risking exposure of sensitive, classified information.

Research describe here has been supported in part by National Science Foundation grant CNS-0831442 CT-M: Theory and Practice of Accountable Systems, IARPA Policy Assurance for Private Information Retrieval grant FA8750-07-2-0031 and the Department of Homeland Security Accountable Information Systems grant N66001-12-C-0082. However, the views expressed here are solely the author's and do not imply an endorsement of the views expressed here by those agencies.