Computer Science and Artificial Intelligence Laboratory

Technical Report

MIT-CSAIL-TR-2006-007 January 27, 2006

Transparent Accountable Data Mining: New Strategies for Privacy Protection

Daniel J. Weitzner, Harold Abelson, Tim Berners-Lee, Chris Hanson, James Hendler, Lalana Kagal, Deborah L. McGuinness, Gerald Jay Sussman, and K. Krasnow Waterman

massachusetts institute of technology, cambridge, ma 02139 usa — www.csail.mit.edu

Transparent Accountable Data Mining: New Strategies for Privacy Protection

[1]Daniel J. Weitzner, [1]Harold Abelson, [1]Tim Berners-Lee, [1]Chris Hanson, [3]James Hendler, [1]Lalana Kagal, [2]Deborah L. McGuinness, [1]Gerald Jay Sussman, [1]K. Krasnow Waterman

[1]Massachussetts Institute of Technology, Computer Science and Artificial Intelligence Laboratory, 32 Vassar St., Cambridge, MA USA
[2]Knowledge Systems Laboratory, Stanford University, Stanford, CA USA
[3]University of Maryland, MIND Lab, 8400 Baltimore Ave., College Park, MD USA

djweitzner@csail.mit.edu, hal@mit.edu, timbl@csail.mit.edu, cph@csail.mit.edu, hendler@cs.umd.edu, lkagal@csail.mit.edu, dlm@ksl.stanford.edu, gjs@mit.edu, kkw@mit.edu

Abstract

Attempts to address issues of personal privacy in a world of computerized databases and information networks -- from security technology to data protection regulation to Fourth Amendment law jurisprudence -- typically proceed from the perspective of controlling or preventing access to information. We argue that this perspective has become inadequate and obsolete, overtaken by the ease of sharing and copying data and of aggregating and searching across multiple data bases, to reveal private information from public sources. To replace this obsolete framework, we propose that issues of privacy protection currently viewed in terms of data *access* be re-conceptualized in terms of data *use*. From a technology perspective, this requires supplementing legal and technical mechanisms for access control with new mechanisms for transparency and accountability of data use. In this paper, we present a technology infrastructure -- the Policy Aware Web -- that supports transparent and accountable data use on the World Wide Web, and elements of a new legal and regulatory regime that supports privacy through provable accountability to usage rules rather than merely data access restrictions.

I. Introduction

Information systems upon which we depend are becoming ever more complex and decentralized. While this makes their power and flexibility grow, it also raises substantial concern about the potential for privacy intrusion and other abuses. Understanding how to incorporate transparency and accountability into decentralized information systems will be critical in helping society to manage the privacy risks that accrue from the explosive progress in communications, storage, and search technology. A prime example of a growing, decentralized information system is the World Wide Web, recently augmented with structured data capabilities and enhanced reasoning power. As the Web gets better and better at storing and manipulating structured data it will become more like a vast global spreadsheet or database, than merely a medium for easy exchange and discovery of documents. Technologies such as XML, Web Services, grids, and the Semantic Web all contribute to this transformation of the Web. While this added structure increases inferencing power, it also leads to the need for far greater transparency and accountability of the inferencing process. By *transparency* we mean that the history of data manipulations and inferences is maintained and can be examined by authorized parties (who may be the general public). By *accountability* we mean that one can check whether the policies that govern data manipulations and inferences were in fact adhered to. Transparency in inferencing systems enables users to have a clear view into the logical and factual bases for the inferences presented by the system. Accountability in inferencing enables users or third parties to assess whether or not the inferences presented comply with the rules and policies applicable to the legal, regulatory or other context in which the inference is relied upon.

Today, when an individual or an enterprise uses a single, self-contained set of data and applications, the controls necessary to assure accuracy and contextualize the results of queries or other analyses are available and generally well understood. But as we leave the well-bounded world of enterprise databases and enter the open, unbounded world of the Web, data users need a new class of tools to verify that the results they see are based on data that is from trustworthy sources and is used according to agreed upon institutional and legal requirements. Hence, we must develop technical, legal and policy foundations for transparency and accountability of large-scale aggregation and inferencing across heterogeneous data sources. We can expect a wide range of legal and regulatory requirements on inferencing systems, and some requirements may well overlap or contradict others. This expected diversity of rulesets makes in all the more important to have one common technical framework for managing accountability to rules.

Such transparency and accountability will be important in a variety of cases: for compliance with financial regulations [SOX] and new security and privacy rules for health care data [HIPAA]. Finance and health are just two areas in which the higher quality data management practices are seen as important in connect with greater

reliance on complex information systems. In the most general case, we will trust inferences only when we have a transparent view into their antecedents and will use them appropriately only when we know that we may be held accountable for misuse. A wide range of public and private sector data mining and inferencing applications will benefit from the transparency and accountability mechanisms described here [JoCrPa04]. One particularly vivid example of this need is the case of government use of large-scale data mining systems for law enforcement and national security purposes.

Transparency and accountability are important features of a larger architectural project to make Web more 'policy aware'. *Policy awareness* is a property of the Semantic Web that will provide users with accessible and understandable views of the policies associated with resources, enable agents to act in response to rules on a user's behalf, thereby making compliance with stated rules easier, and afford a greater opportunity for accountability when rules are intentionally or accidentally broken. [WHBC05]

Our exploration of transparency and accountability as privacy protection mechanisms begins with elaboration of a government data mining privacy scenario drawn from the actual debate over the design and regulation of the proposed airline passenger screening system in the United States. This simple scenario will illustrate the privacy problems posed by large-scale profiling of individuals and then show how increased transparency and accountability to a clearly defined set of data usage rules can support fundamental privacy values. Based on our implementation experience with the scenario described here, we propose a technical architecture that will enable privacy compliance. For this purpose, we draw upon the Semantic Web technology which is laying the foundation for tagging and classifying data at Web scale, and we combine this with technology for automated deduction and justification of conclusions across large-scale databases and multiple reasoning systems.

The fundamental technical challenge that must be addressed in order to provide transparency and accountability for reasoning on the Semantic Web is rooted in the open, decentralized architecture of the Web itself. The Semantic Web [BLHL01] is an enhancement of the current Web to allow machine-processable data to span application boundaries in the same way that human-readable documents do currently. The goal of the Semantic Web is as broad as that of the Web: to be a universal medium for data. It is envisaged eventually to smoothly interconnect personal information management, enterprise application integration, and the global sharing of commercial, scientific and cultural data. Introducing transparency into the reasoning occurring over the Semantic Web requires innovative techniques that account for the open, decentralized architecture of the Web.

Beyond the basic architecture of the Web, four more general trends in the use of information should encourage privacy-sensitive system designers to rethink their approach to privacy protection: first, the gradual demise of stove-pipe applications in favor of enterprise-wide data integration; second, the rapidly declining cost of web-scale query; and third, the rapid spread of sensor networks in both public and private settings. Fourth, the cost of data storage is becoming cheaper and cheaper to the point that is often less expense to just keep all data rather than figure out which information to discard and which to retain. No doubt, there is a fixed cost associated with operation of data storage facilities, but with the rapidly declining cost of disk storage, the cost per data element is approaching zero.

Current technical investigations of the impact of data mining on privacy have generally focused on limiting access to data at the point of collection or storage. As we will discuss, much effort has been put into the application of cryptographic and statistical techniques to construct finely tuned access-limiting mechanisms. Yet for all this emphasis on access restriction, the reality is that the Web is making it increasingly difficult to limit access to data, while at the same time making it increasingly easy to aggregate data from multiple information sources, and to do searching and inferencing based on these aggregations. In the long run, access restriction alone cannot suffice neither to protect privacy nor to ensure reliable conclusions. It must be augmented by attention to increased transparency and accountability for the inferencing and aggregation process itself.

From a public policy perspective, the emphasis on usage limitation as opposed collection limitation is unconventional and perhaps controversial. Following the description of the proposed TAMI architecture, we will show how basing regulatory schemes governing privacy and data mining on transparency can serve as a basis for achieving basic privacy goals. We will explore analogues to current-day Fourth Amendment protections that consider not only access to information, but also the ways in which diverse information sources are aggregated and the ways in which implications are drawn.

II. Illustrating the Data Mining Privacy Challenge

As a law enforcement and national security tool, data mining holds out the promise of being an important new component of criminal investigation and terrorism prevention, but raises at the same time a whole new category of privacy challenges [Mark02]. The power of data mining technology lies in its potential to bring to light non-obvious investigation targets or identify terrorist threats through inferences drawn on decentralized data sets spread around the Web, around the world. This qualitative expansion in inferencing power is viewed as important to

keep pace with new security threats, but also puts an unprecedented level of intrusive power in the hands of government.

A. Scenario: Rules for Usage of Passenger Profiling Information

It is possible to develop general purpose transparency mechanisms for Semantic Web reasoning and then apply those tools in data mining environments. At the heart of the debate over the design of the proposed airline passenger screening systems (CAPPS, CAPPS II, and now Secure Flight) is the question of whether data collected in the course of assessing security risks can then be used for other law enforcement purposes. We illustrate (Fig.1) some of the unanswered privacy problems associated with use of data mining for law enforcement and/or national security purposes. We then describe how the use of truth maintenance systems and proof checking techniques can assure both transparency of the facts behind decision making and accountability for adhering to appropriate use limitations on data as it flows across previously well-established institutional boundaries. With this Policy Aware architecture in place *and* a clear set of legal rules in place, it is possible to address the key privacy protection requirements of government data mining.

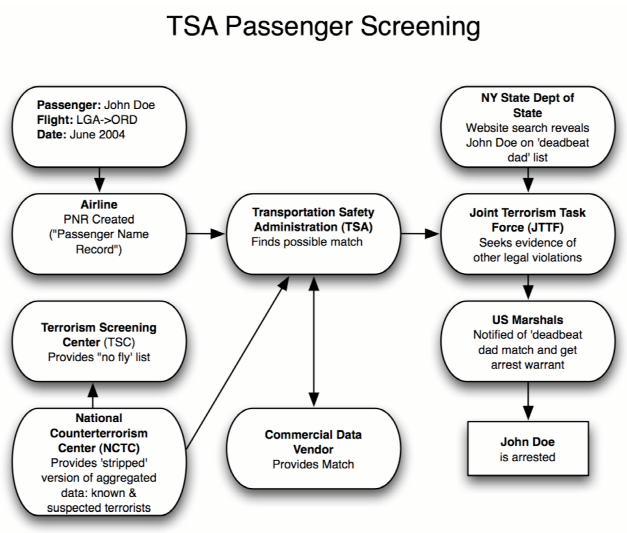


Figure 1

In this scenario, a traveler named John Doe from New York boards a flight in New York and sets in motion a chain of inferences (some of which are factually incorrect and some of which are reached in violation of rules) that generates a series of adverse consequences for him.

Doe traveled on a flight from New York to Chicago in June 2004. Under the Department of Homeland Security (DHS), Transportation Security Administration's (TSA's) test of its Secure Flight program, it has accumulated his Passenger Name Record (PNR) from the airline: data from a commercial data vendor including full name, address, date of birth, and gender; and a "no fly" list from the Terrorism Screening Center of people known or reasonably suspected of being associated with terrorism. [SF2005]

Mr. Doe is matched to the "no fly" list, but it is unclear whether he actually is the person on the list or just one of several people who share the name and birth date. Due to clerical error, the PNR data does not contain a complete address, only the city and state (New York, New York). Because "John Doe" is such a common name, the commercial vendor associates the name with several addresses. It is unclear whether the addresses are associated with more than one John Doe with the same birthday, or if the traveler John Doe has lived at multiple addresses. Doe has long since taken the flight, so he is not physically present at an airport where a TSA employee can ask follow up questions.

The TSA employee reviewing the test results is concerned about the possibility that the person could be the terrorist identified by the TSC. Under the existing Routine Use notice for Secure Flight testing, he notifies the Joint Terrorism Task Force (JTTF) in New York. The agents there agree that they would like to know more about Doe. They research John Doe of 123 Main Street and find no evidence to support the idea that he is associated with terrorism. However, while researching him, the agents match his name to a large outstanding child support obligation through a New York state website. There is a federal "deadbeat dad" law providing criminal penalties for this. The JTTF gets the details of the New York state case and an arrest warrant is obtained. John Doe is found and arrested.

B. Privacy failure modes

There are at least privacy three failures in this scenario that could be addressed by greater transparency and accountability.

First, it happens that the John Doe who was on the plane was not the John Doe who lives on 132 Main St. With transparency tools in place, he could have been given an easy option to verify whether the "proof" that resulted in heightened suspicion was actually based on factually true antecedents; he, of course, could have then shown this was not the case. There may be security reasons why some of these antecedents would have to be obscured, but reasonable transparency into the proof tree used by TSA could have saved him the intrusion of the screening, and saved TSA the unnecessary expense. As will be discussed

more below, this requires a transparent reasoning system that maintains the proof tree for evaluation when needed.

Second, under current regulations, TSA was authorized to share information about a person with another agency only if there was a reasonable belief that the person is related to terrorism. Without transparent reasoning and accountability measures, a well-meaning TSA agent might not even know that passing the John Doe information was a violation. A TSA agent who was aware that such sharing is wrong might think twice before doing so if s/he knows that accountability mechanisms would catch the unauthorized action.

Third, the JTTF is permitted to use the information received for a purpose only in a manner consistent with the purpose for which the data was collected. The information about John Doe was collected to identify and pursue terrorists. The JTTF members could be wholly unaware that using the information in a purely domestic, criminal context such as a deadbeat dad investigation was inappropriate. With transparent reasoning capabilities in place, the system could highlight this rule violation.

The scenario described here, even though it is vastly simpler than actual homeland security data mining applications, demonstrates the real challenges of preserving privacy and monitoring government conduct in the web-like, decentralized law enforcement information network that is currently coming into being.

C. Privacy requirements for data mining

We have identified three distinct classes of rule violations, as measured by either current data handling rules or laws that we would expect to be put into place:

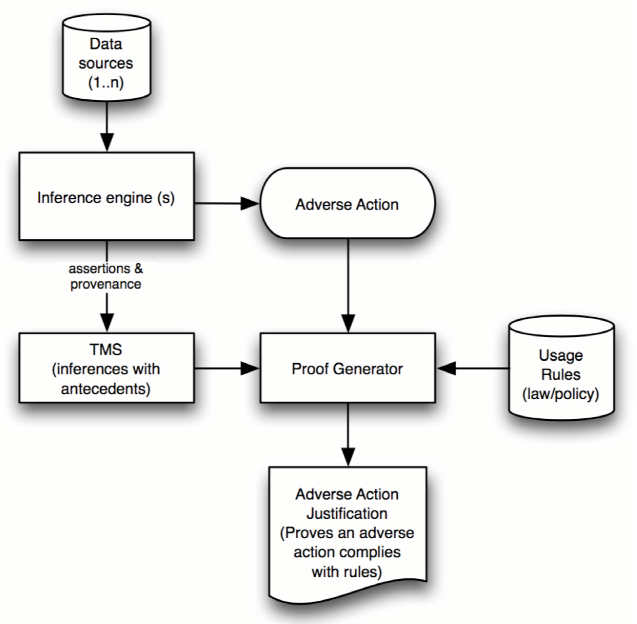
1. Adverse actions premised on factually incorrect antecedents
2. Impermissible sharing of data beyond the collecting organization
3. Adverse actions premised on inferences from data where the data, while factually correct and properly in possession of the user, is *used* for an impermissible purpose.

The first two cases can be handled with audit and verification mechanisms of the sort that are technically well understood and commercially available today. However, the third problem requires *a posteriori* assessment of rules compliance –i.e., accountability. It is only when the data is actually used (long after collection) for an impermissible purpose that the rule violation can be discovered. In logical terms, the conclusion of the proof relies upon antecedents that *logically* support the conclusion but are not *legally* permitted to be used to support such a conclusion.

III. Basic Architecture for Transparent, Accountable Data Mining

A. TAMI Architecture

In order to meet the above requirements, we propose an information architecture consisting of general-purpose inferencing components connected in a manner (Figure 2) that provides transparency of inferencing steps and accountability to rules.



TAMI Functional Architecture (Figure 2)

The transparency and accountability architecture depends upon three components:

- **Inferencing Engine(s):** support analysis of data available and assesses compliance with relevant rules
- **Truth Maintenance System**: a persistent store fed by the inference engine(s) consisting of proof antecedents as well as data provenance, used to assess reliability of inferences and to record justifications for proof antecedents developed in the course of an investigation.
- **Proof Generator:** constructs proofs that critical transitions and adverse uses of personal information are justified by facts and permissible under applicable rules

The inference engine provides assistance to the government investigator or analyst in identifying

suspicious profiles in those data sets accessible for this purpose. This data would then be processed through an inferencing engine (we use the cwm engine [CWM00] in this case) that provides investigative results.

In addition to these investigatory inferences, a record of the inferences and their justifications will be stored in the Truth Maintenance System (TMS) [Do87][BrKa03] . The TMS combined with a proof generator allows anyone with access to the system to determine whether or not the personal information in the system is being used in compliance with relevant rules and consistent with known facts. At critical stages of the investigation, such as sharing of information across agency boundaries or use of information to support an adverse inference (secondary screening, criminal indictment, arrest, etc.), the proof generator will attempt to construct a proof that the use proposed for the data at that transition point is appropriate. The proof generator would be able to draw on information collected in the TMS and bring to bear the relevant rule sets.

Applying this general framework to the screening scenario, we can see how it addresses each of the three failure modes identified.

1. Identify factually incorrect antecedents

Based on the investigative inference that Mr. Doe is on a terrorist watch list, we can expect that a government screening system will cause him to be stopped at the airport security checkpoint in order to subject him to secondary screening. An appropriate authority could use a transparent reasoning system to factor out classified information and test the antecedents that led to the stop instruction. The system could identify that one antecedent of the proof that he should be stopped is the factual assertion that he lives at 132 Main St. Without necessarily revealing the content of this antecedent (the address), system could then ask Mr. Doe for documentation of his address. When Mr. Doe presents valid documentation of a different address, the proof would be re-evaluated with the result, in this case, that Doe can board the plane, sparing himself the intrusion and saving TSA the unnecessary resources.

2.* Assess compliance with information sharing rules before data *transfer

As an investigation proceeds, the inferences drawn may lead an analyst to believe that information ought to be shared with other law enforcement or national security agencies. Just as inference engines and truth maintenance systems working together can help the analyst assess the reliability of information developed in the course of an investigation, so too can these mechanisms provide an investigator in one agency guidance on the question of whether information may permissibly shared with another agency. In our scenario, the rule is that information sharing is allowed only when there the investigator has reason to believe that the subject of the investigation is related to a terrorist threat.

Upon initiating a transfer of information, the system could seek to generate a proof that such sharing is permitted. If the truth maintenance system contains no basis for such a proof, then a warning could prevent sharing or alert the investigator about to share the data to a potential rule violation.

3. Check that adverse actions are consistent with information usage rules

The configuration of inference engines, truth maintenance systems and proof checking have a unique role to play in providing accountability to rules when an adverse consequence is proposed as result of the use of personal information in a profiling process. An accountable profiling system will be able to bind a proof of rule compliance together with a conclusion justifying an adverse consequence. A proof generator is therefore needed in order to assess whether there are sufficient assertions stored in the TMS to justify whatever consequence is proposed. In our hypothetical scenario, the proof generator would attempt to construct a proof that the information from the TSA is properly used to justify the arrest of Mr. Doe as a deadbeat Dad. Though he might actually be guilty of that crime, the data usage rules clearly prevent passenger screening data from being used for this purpose.

To the extent that the TAMI architecture is able to close the gap left by these three privacy failure modes, we can see the importance of having such proof-based transparency and accountability systems in place where large-scale government data mining is contemplated.

B. Current Implementation Status

Our initial work implementing the TAMI architecture has been addressing the challenges of communicating seamlessly between the legal, logical, and semantic web structures. Using current United States Government efforts as a guide, we presume that the historical log of data collection, analysis, and transfer, as well as case activities, will exist in XML. Where possible, we used the recently released National Information Exchange Model (NIEM) [NIEM], the joint Department of Justice and Department of Homeland Security XML interchange format for law enforcement investigative data. Building on our hypothetical, we created a fictional transaction log. This version assumes that the transaction is traced back through multiple agencies' records and that the relevant items were concatenated into a single file. Then we created a "cleansed" version (sample at Figure 3), which assumes that some system reorganized the data into a more organized, readable format.

```
<event id="flight-test-search-1">
  <name>Secure Flight Test</name>
  <type>search</type>
  <xsd:date/>
  <search-query ref="query-1"/>
  <search-result ref="result-1-1"/>
  <search-result ref="result-1-2"/>
  <search-result ref="result-1-3"/>
</event>

<search-query id="query-1">
<terms>
<u:PersonGivenName>John</u:PersonGivenName>
<u:PersonMiddleName>Henry</u:PersonMiddleName>
<u:PersonSurName>Doe</u:PersonSurName>
<u:PersonBirthDate>1975-08-
24</u:PersonBirthDate>
</terms>

</search-query>
<search-result id="result-1-1">
<search-query ref="query-1"/>
<source ref="TSDB"/>

<items>
<u:PersonGivenName>John</u:PersonGivenName>
<u:PersonMiddleName>Henry</u:PersonMiddleName>
<u:PersonSurName>Doe</u:PersonSurName>
</items>
</search-result>
```

Sample transaction log in XML using NIEM (Figure 3)

We can use XSL Transformations to automatically convert the XML transactional data into RDF [RDF]. We use the "Notation 3" [N3] notation for serializing RDF, producing results similar to the class an instance definitions shown in Figures 4 and 5.

```
:Database a rdfs:Class.
:owner a rdf:Property; rdfs:domain :Database;
rdfs:range :Organization.

:DataRecord a rdfs:Class.
:PassengerNameRecord a rdfs:Class;
rdfs:subClassOf :DataRecord.

:source a rdf:Property; rdfs:domain
:DataRecord; rdfs:range :Database.
:date a rdf:Property; rdfs:domain :DataRecord;
rdfs:range xsd:Date.

:passenger a rdf:Property; rdfs:domain
:PassengerNameRecord; rdfs:range :Person.
:flight a rdf:Property; rdfs:domain :
PassengerNameRecord; rdfs:range :Flight.

:Flight a rdfs:Class.

:date a rdf:Property; rdfs:domain :Flight;
rdfs:range xsd:Date.

:number a rdf:Property; rdfs:domain :Flight;
rdfs:range :Literal.
:origin a rdf:Property; rdfs:domain :Flight;
rdfs:range air:Iata.
:destination a rdf:Property; rdfs:domain
:Flight; rdfs:range air:Iata.
```

Sample Classes in RDF serialized in N3 (Figure 4)

```
:pnr-1 a :PassengerNameRecord;
   :source :AA-PNR;
   :date 2004-06-14;
   :passenger
     [:name
        [:personGivenName "John",
         :personMiddleName "Henry",
         :personSurName "Doe"];
      :birthDate 1975-08-24];
   :flight
     [:number "723",
      :date 2004-06-14,
      :origin :LGA,
      :destination :ORD].

:flight-test-search-1 a :Search;
:date 2005-09-12;
:name "Secure Flight Test";
:query [:aboutPerson
         [:name
           [:personGivenName "John",
            :personMiddleName "Henry",
            :personSurName "Doe"];
          :birthDate 1975-08-24]].
```

Sample RDF Instances serialized in N3 (Figure 5)

We will be expressing laws in N3 logic [CWM] over the transactional data in RDF. This requires us to build common understanding about how to convert law to rules in N3. For example, the "Deadbeat Dad" statute includes as the part of the definition of "a failure to pay legal child support obligation" offense the condition that a person:

> willfully fails to pay a support obligation with respect to a child who resides in another State, if such obligation has remained unpaid for a period longer than 1 year, or is greater than $5,000. 18 U.S.C. §228(a)(1)

This is expressed in N3 logic as (Figure 6):

```
@keywords a, is, of, this.
@prefix log: <http://www.w3.org/2000/10/swap/log#>.
@prefix math: <http://www.w3.org/2000/10/swap/math#> .
@prefix string:
   <http://www.w3.org/2000/10/swap/string#>.

@prefix geo:
   <http://opencyc.sourceforge.net/daml/cyc.daml#>.
```

```
@prefix usps:
   <http://www.w3.org/2000/10/swap/pim/usps#>.

{ ?X a Person.
   ?X outstandingObligation ?ChildSupport.
   ?ChildSupport a ChildSupportOutstandingObligation.
   ?ChildSupport value ?Amt.
   ?Amt math:greaterThan 1000.
   ?ChildSupport obligee ?Y.
   ?Y a Child.
   ?X residence [ geo:inRegion [ usps:stateAbbr
     ?XState ]].
   ?Y residence [ geo:inRegion [ usps:stateAbbr
     ?YState ]].
   ?XState string:notEqualIgnoringCase ?YState. }
=> { ?X a :DeadbeatDad }.
```

"Deadbeat Dad" Law in N3
(Figure 6)

This is a short statute, that uses near mathematical logic, Translating this statute into N3 has been an important first step in determining that we could in fact use N3 to represent laws. Implementation of this rule will help us address our first failure model: identifying factually inaccurate antecedents.

We have also confirmed that the cwm reasoning engine[CWM] can be used as a logic system for this application. In our "Deadbeat Dad" example, we created a simple set of facts in N3 (Figure 7), a filter to return the positive and negative results, and were able to fire the rules (Figure 6) successfully.

```
# Facts
Joe a Person.
Sue a Child.
:05-CIV-NY-223 a OutstandingObligation.
Joe outstandingObligation :05-CIV-NY-223.
:05-CIV-NY-223 a
   ChildSupportOutstandingObligation.
:05-CIV-NY-223 obligee Sue.
:05-CIV-NY-223 value 1500.

Joe residence
   [ geo:inRegion [ usps:stateAbbr "NY"]].
Sue residence
   [ geo:inRegion [ usps:stateAbbr "MA"]].
```

Facts to Run "Deadbeat Dad" Against
(Figure 7)

Our initial efforts at representation of a more complex law (The Privacy Act, 5 U.S.C. § 552a) reveal the challenge of working with multiple logical structures and requires the ability to reach out to other sources to complete the firing of the rule For example, the Privacy Act has more than 20 separate rules that set the criteria for what data an agency can collect, what information the agency must disclose about the sources of the data and what information the agency must disclose about its decisions to share data. Representation of laws with this kind of complexity is a prerequisite for addressing the second failure mode, assessing whether an agency is permitted to have access to or ownership of data.

In keeping with the architecture of the Semantic Web we use Uniform Resource Identifiers (URI) to identity our rules and have tentatively identified a naming convention for them. The Legal Institute of Cornell Law School has already provided URIs for laws, to the subsection level, allowing us to provide specific source references. We are expecting to use a Truth Maintenance System as the storage mechanism for our proofs and, possibly, as an alternate deductive reasoner. We have produced one sample, using the "Deadbeat Dad" rules in AMORD [KDSS77]. We expect to continue discussions regarding monotonic, non-monotonic, or other logic schemes as we expand our samples.

We will register the reasoning engine we use, cwm, in the Inference Web [IW][McP04]. We will use Inference Web to browse conclusions produced by our reasoner. End users can inspect how conclusions were deduced, what sources were relied on, and any provenance information about the sources such as date, source author, etc. Inference Web may also be used to abstract the explanation and meta information in multiple formats.

Once cwm is able to generate PML [PMF05], , we will define policies over these proof trees to confirm that the antecedents of every node in a proof tree were collected and used in accordance to the Privacy Act.

IV. Transparency and Accountability in the Current Privacy Policy Debate

We have shown that large scale data mining poses novel privacy challenges which require response. However, our efforts to structure laws and develop technologies with sensitivity for privacy values should seek guidance from the nearly century-long interplay between ever-growing surveillance capabilities of new technologies and fundamental privacy principles. Historically, we learn that as electronic communications have become more sophisticated and more ubiquitous, communications privacy law has responded to the advance in law enforcement needs *and* privacy threats by tying the growth in surveillance capabilities to gradually expanding privacy protections that kept pace with new intrusion powers. Over the last hundred years in the United States and elsewhere around the world, privacy protections were extended to voice telephone calls, then email, then transactional records, and other communications-related information [De97]. Web-scale inferencing that powers data mining is only the latest in the series of technology advances that

demands new privacy protection alongside intrusive surveillance powers [Hsrpt86].

The inherent complexity of data mining dictates, as our scenario shows, that privacy values will not be protected merely by controlled access to personal information in the way that wiretapping laws could simply grant or deny access to a telephone conversation. We will have to supplement *a priori* access control with *a posteriori* accountability to rules. As the passenger screening scenario demonstrates, privacy protection will require both the ability to assure that adverse actions are premised on factually correct antecedents, and that the adverse conclusions are logically grounded in permissible uses of personal information. As the conclusions are reached and acted upon long after the information supporting those conclusions were collected, we obviously cannot rely upon *a priori* control mechanisms operating only at the time of collection. Rather, full accountability to privacy rules cannot be achieved without the *a posteriori* proof techniques we have described here.

Transparency and accountability mechanisms are a vital part of privacy protection going forward because we expect continued expansion in the depth and breadth of data available both to the government and the private sector. The great power of data mining to reveal intimate details about individuals has yet to be matched with either legal or technical measures that balance its impact with privacy requirements [CDT03]. What's more, there are proposals to expand law enforcement data analysis powers even further. In calling for the creation of a nationwide network to respond to threat of terrorism, a Markle Foundation Task Force explains that an open, decentralized Web-like architecture is really the only design strategy that could possibly succeed in linking that many disparate entities in law enforcement, homeland security, intelligence, and defense with a role to play. In addition to the twenty-two federal agencies now under the DHS umbrella, the following organizations must be integrated into a single, coordinated information sharing environment:

18 federal agencies in the US cabinet
17,784 State & Local law enforcement agencies
30,020 Fire departments
5,801 Hospitals
1,700 Private critical infrastructure
[BJS2000][Pa2004]

In such a far-flung and heterogeneous environment, both collection and analysis of data must "occur at multiple nodes, rather than only in a few centralized locations" [Mark03]. Reliance on Web architecture as a model for sharing, analyzing, and managing this data is appropriate not because of any desire to make all of this data public (as much of the Web is) but because institutions have learned that the decentralized addressing model of the Web has been uniquely successful in enabling large-scale coordination of data both inside and outside enterprise boundaries.

How much larger that universe of data grows and how quickly this happens is a matter for public policy makers to decide in an open, democratic process. As technology designers, however, we can provide information infrastructure that help society be more certain that data mining power is used only in legally-approved ways, and that the data which may give rise to adverse consequences for individuals is based on inferences that are derived from accurate data. We can meet these goals by making sure that the architecture of new Web technologies provides transparency into the inferencing mechanisms and creates technical means for assuring that government data mining efforts are accountable for improper use of data.

An alternative to privacy protecting data mining algorithms

Our proposal to rely on transparency and accountability as privacy protection mechanisms stands in contrast to other efforts to engineer privacy protection into information systems. Recently, much work has been done on distributed database systems with secure private computation algorithms (SPCA) [GoMi82] as a means of protecting privacy [BFSW04]. Privacy-preserving data mining algorithms [LiPi02] have shown that it is possible to constrain query power based on some predefined measure of how much information the requestor is entitled to have and some quantified notion of privacy [EGS03]. While such systems may well have their place in some privacy applications, it has not yet been demonstrated that they can be successfully deployed at the scale required to meet privacy requirements for either large scale private sector or government data mining. What's more, the ability to constrain queries in this manner depends on a mathematically-expressible definition of privacy describing the quantitative limits on how much information the government can have [AgSr00]. As we have shown, compliance with privacy rules can often depend on factual circumstances only manifest *after* a given query has been made, so it is simply impossible to rely on control over query (data collection rules) alone to protect privacy. Furthermore, it will not always be possible to articulate a computable definition of privacy. In many cases, privacy laws rely on some judgment of whether one set of facts 'reasonably' justifies access to some larger set of information, as is the case with a “probable cause” requirement for electronic surveillance. Finally, while SPCA can enable control of the scope of queries within the bounds of a given information system, data may leak out of systems instrumented with SPCA through a variety of channels, not subject to control of the query control mechanisms.

We believe that reliance on secure, private computation algorithms both under-emphasize the vital need for

transparency into the use of data mining, and also may result in over-constraining the use of data mining capability to the detriment of law enforcement needs. Even if such privacy-preserving data mining techniques prove to be practical, they are unlikely to provide sufficient public assurance that government inferences conform to legal restrictions. They also do not address the need to provide citizens the certainty that adverse government action is based on factually accurate data. In sum, while privacy-preserving data mining techniques are certainly necessary in some contexts, they are not sufficient privacy protection without the transparency and accountability.

Toward a public policy agenda based on transparency and accountability

Transparency and accountability technologies are necessary, but certainly not sufficient for privacy protection in an age of large scale public and private sector data mining. Our Policy Aware Web infrastructure can provide meaningful privacy protection through transparency and accountability only if social conventions and legal requirements make such mechanisms available and effective. While it is beyond the scope of this paper to develop detailed public proposals, we believe that policy aware systems bring added focus to policy questions regarding data mining privacy. In order to realize the promise of transparency and accountability in support of privacy values, the legal system will have to address questions such as these:

- What degree of transparency rights (also known as 'access rights' in privacy law) should those subject to data mining have?
- What will be the mechanism for correction of data found to be incorrect?
- Will there be legal recourse in the event agencies rely on incorrect information after the error has been pointed out by the data subject?

Accountability mechanisms hold significant promise, but only meaningful if the legal rules against which data miners are held accountable are properly reflective of privacy values. Rules are needed to address questions such as:

- Under what circumstances, if ever, can inferences generated in one type of profiling system (anti-terrorism passenger screening, for example) be used to further criminal investigations?
- If data mining results can be shared across the national security/domestic criminal investigation "wall", is this true in all cases or only for certain classes of crimes?
- If data mining is used in a criminal investigation, can those results be applied to any other type of crime? For example, should someone under suspicion of late tax payment also be subject to checks for unpaid parking tickets or expired drivers license.

The Policy Aware systems we have described have the ability to deal with a wide range of rules in the above categories, but the rules, whatever they are, must be specific enough provide real transparency and accountability.

V. Conclusion

Our goal is to develop technical and legal design strategies for increasing the transparency of complex inferences across the Semantic Web and data mining environments. We believe that transparent reasoning will be important for a variety of applications on the Web of the future, including compliance with laws and assessing the trustworthiness of conclusions presented by reasoning agents such as search engines. Our particular focus is on using transparent inferencing to increase accountability for compliance with privacy laws. We also expect that this technical research will provide important guidance to policy makers who are considering how to fashion laws to address privacy challenges raised by data mining in both private sector and homeland security contexts.

Acknowledgements

Work conducted at the MIT CSAIL Decentralized Information Group with support from National Science Foundation grants: the Transparent Accountable Data Mining Initiative (award #0524481) and Policy Aware Web project (award #0427275).

References

[AgSr00] Agrawal D. and Srikant, R. Privacy preserving datamining, Proc 2000 ACM SIGMOD Conference on Management of Data, 2000, 439-450

[BFSW04] D, Boneh, J. Feigenbaum, A. Silberschatz, R. Wright, PORTIA: Privacy, Obligations, and Rights in Technologies of Information Assessment, Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 27, pp. 10-18 (2004).

[BLHL01] Berners-Lee, T., Hendler, J. and Lassila, O. The Semantic Web: When the Internet gets smart, Scientific American, May 2001.

[BJS2000] http://www.ojp.usdoj.gov/bjs/lawenf.htm (last visited 25 October 2005)

[BrKa03] Jeen Broekstra and Arjohn Kampman. Inferencing and Truth Maintenance in RDF Schema: Exploring a naive practical approach. In Workshop on Practical and Scalable Semantic Systems (PSSS), Sanibel Island, FL, 2003.

[CDT03] CDT Report - Privacy's Gap: The Largely Non-Existent Legal Framework for Government Mining of Commercial Data, May 28, 2003. http://www.cdt.org/security/usapatriot/030528cdt.pdf

[CWM] Tim Berners-Lee and Dan Connolly and Eric Prud'homeaux and Yosi Scharf, Experience with N3 rules, W3C Rules language Workshop, 2005.

[CWM00] Berners-Lee, T., CWM – A general purpose data processor for the Semantic Web, 2000. http://www.w3.org/2000/10/swap/doc/cwm.html

[De97] J. Dempsey, "Communications Privacy In The Digital Age: Revitalizing The Federal Wiretap Laws To Enhance Privacy," Albany Law Journal of Science & Technology, 1997. http://www.cdt.org/publications/lawreview/1997albany.shtml

[Do87] J. Doyle. A Truth Maintenance System. In Readings in Nonmonotonic Reasoning, pages 259–279. Morgan Kaufmann Publishers, San Francisco, CA, USA, 1987.

[EGS03] A. Evfimievski, J. Gehrke, and R. Srikant, "Limiting Privacy Breaches in Privacy Preserving Data Mining," in Proceedings of the 22nd Symposium on Principles of Database Systems, ACM Press, New York, 2003, pp. 211–222.

[GoMi82] S. Goldwasser, S.Micali, Probabilistic encryption & how to play mental poker keeping secret all partial information, Proceedings of the fourteenth annual ACM symposium on Theory of computing, pp. 365-377, 1982, ACM Press.

[Hsrpt86] United States House of Representatives, Juciary Committee Report on the Electronic Communications Privacy Act of 1986 (House Report 99-647).

[HIPAA] Health Insurance Portability and Accountability Act of 1996 (Pub. L. 104-191).

[IW] Deborah L. McGuinness and Paulo Pinheiro da Silva. Explaining Answers from the Semantic Web: The Inference Web Approach. Journal of Web Semantics. Vol.1 No.4., pages 397-413, October 2004.

[JoCrPa04] D. Johnson, S. Crawford, J Palfry, The Accountable Internet: Peer Production of Internet Governance, 9 Virginia Journal of Law and Technology 9 (2004)

[KFJ03] L. Kagal, T. Finin, A. Joshi, "A Policy Based Approach to Security for the Semantic Web", In Proceedings, 2nd International Semantic Web Conference (ISWC2003), September 2003.

[Ka04] L. Kagal, "A Policy-Based Approach to Governing Autonomous Behavior in Distributed Environments", Phd Thesis, University of Maryland Baltimore County, November 2004.

[KDSS77] J. de Kleer, J. Doyle, G. L. Steele Jr., and G. J. Sussman. *Amord: Explicit control of reasoning*. In Proceedings of the ACM Symposium on Artificial Intelligence and Programming Languages, pages 116–125, 1977.

[Ko92] R. Kowalski, "Legislation as Logic Programs," In: Logic Programming in Action (eds. G. Comyn , N. E. Fuchs, M. J. Ratcliffe), Springer- Verlag, pages 203-230 (1992).

[LiPi02] Y. Lindell and B. Pinkas, "Privacy preserving data mining," J. of Cryptology, 15:177-206, 2002.

[McFiMc03] McCool, R.; Fikes, R.; & McGuinness, D. Semantic Web Tools for Enhanced Authoring. KSL, 2003. http://www.ksl.stanford.edu/KSL_Abstracts/KSL-03-07.html

[McP04] Deborah L. McGuinness and Paulo Pinheiro da Silva. Explaining Answers from the Semantic Web: The Inference Web Approach. Journal of WebSemantics. Vol.1 No.4., pages 397-413, October 2004.

[Mark02] Protecting America's Freedom in the Information Age. Markle Foundation, 2002. http://www.markle.org/downloadable_assets/nstf_full.pdf

[Mark03] Creating a Trusted Network for Homeland Security: Second Report of the Markle Foundation Task Force, 2003. http://www.markle.org/downloadable_assets/nstf_report2_full_report.pdf

[N3] Notation 3. (working papers online) http://www.w3.org/DesignIssues/Notation3.html

[NIEM] National Information Exchange Model, http://www.niem.gov/

[Pa2004] Guarding America: Security Guards and U.S. Critical Infrastructure Protection. Congressional Research Service (14 November 2004) http://www.fas.org/sgp/crs/RL32670.pdf

[PMF05] Paulo Pinheiro da Silva, Deborah L. McGuinness and Richard Fikes. A Proof Markup Language for Semantic Web Services. Information Systems, 2005.

[RDF] RDF Primer, http://www.w3.org/TR/rdf-primer/

[PiMcMc03] P. Pinheiro da Silva, D. L. McGuinness and Rob McCool. Knowledge Provenance Infrastructure. IEEE Data Engineering Bulletin Vol.26 No.4, pages 26-32, December 2003. http://www.ksl.stanford.edu/people/dlm/papers/provenance-abstract.html

[SF2005] Federal Register: June 22, 2005 (Volume 70, Number 119), p. 3619, 3621 (System of Records Notice for Secure Flight, "Categories of Records in the System," subsection (a), describing the acquisition of Passenger Name Records (PNRs) in response to the Transportation Security Administration Order issued November 15, 2004 (69 FR 65625)).

[SOX] Sarbanes-Oxley Act of 2002 (Pub. L. 107-204).

[WHBC05] Weitzner, Hendler, Berners-Lee, Connolly, Creating the Policy-Aware Web: Discretionary, Rules-based Access for the World Wide Web in Elena Ferrari and Bhavani Thuraisingham, editors, Web and Information Security. IOS Press, forthcoming.

[Weit00] Weitzner, D. Testimony before the United States Senate Commerce Committee Hearing on Online Privacy. May 2000