

End-to-End Semantic Accountability

A proposal submitted in response to National Intelligence Community Enterprise Cyber Assurance Program (NICECAP) BAA 06-11-IFKA (211).

Thrust: Accountable Information Flow

Period: Phase I: February 2007 – July 2008
Phase II & Phase III (options)

Cost: Phase I:
Phase II and III: (options)

Institutions: Massachusetts Institute of Technology (DUNS 00-142-5594)
Computer Science and Artificial Intelligence Laboratory
Daniel J. Weitzner, PI (Technical POC)
<djweitzner@csail.mit.edu>
<http://www.w3.org/People/Weitzner.html>
+1 617 253 8036
Hal Abelson, co-PI
Timothy Berners-Lee, co-PI
Chris Hanson, co-PI
Gerald Jay Sussman, co-PI

Rensselaer Polytechnic Institute
James Hendler, co-PI

Yale University
Joan Feigenbaum, co-PI

Program Manager: Walt Tirenin <walt.Tirenin@rl.af.mil>

All work done on this project will be unclassified

Section 2: Innovative Claims and Applications

Key Innovative Claim: Accountable information flow in large-scale, information-sharing environments can be achieved through the novel integration of Semantic Web information-modeling techniques along with Policy Aware Web rules-based access control systems. Using these techniques we will build what we call Accountability Appliances, to be deployed in a decentralized manner throughout complex information spaces such as the World Wide Web such that they can provide system-wide accountability to relevant information usage rules and policies. A properly coordinated system of these accountability appliances will exhibit a property that we call End-to-End Semantic Accountability.

As intelligence requirements drive analysts to seek answers to questions over larger and more heterogeneous information spaces, the challenge of security, privacy law compliance, and assessment of the trustworthiness of results will grow; as much as current information-processing and networking environments lack accountable information flow today, the challenges of security and trustworthiness will grow far harder in the future. To deal with this, we propose to design and implement an information architecture providing "end-to-end semantic accountability" – that is, accountable information flow delivering the ability to both control access to information and to assess compliance with the policies and rules governing data usage. Drawing from our experience in the design and development of the World Wide Web, the Semantic Web, Policy Aware Web systems, reasoning systems, and the application of cryptographic techniques to networking protocols, we will implement a general purpose 'accountability appliance'¹ that can be deployed in a variety of data platforms.

The technical developments proposed here are designed to support a dual social policy goal: a) increased transparency and accountability of information systems, which will lead to b) increased confidence in society that government use of sensitive information, once deemed necessary for protecting national security or criminal law enforcement, will be carried out in a manner that is actually compliant with the rules that policy makers set to protect civil liberties and enable aggressive defense against threats to the country. As we will show, there is an urgent need for transparency and accountability in a variety of information applications that depend upon sensitive, personal information, both in the public and the commercial sector. 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 preventing or controlling access to information. We argue that this perspective has become inadequate and obsolete, overtaken by the effortlessness of sharing and copying data, and the ease of aggregating and searching across multiple data bases to reveal private information from public sources.

To offer meaningful protection of civil liberties, privacy protection, conceptualized today as data access restrictions must be re-conceptualized in terms of data use limitations. From a technology perspective, this requires supplementing legal and technical mechanisms for access control, with new mechanisms for transparency and accountability of data use. We seek to design an information architecture Web that can provide transparent access to the reasoning steps taken in the course of data mining, and establish accountability for use of personal information as measured by compliance with data usage rules.

Accountability Appliances, based on designs that we propose to develop in this project, are essential features in a future information sharing environment. The property of end-to-end semantic accountability that these appliances can provide will assure that information sharing proceeds unimpeded by unnecessary restrictions, but with full knowledge by both the intelligence community and the general public that inappropriate uses of sensitive information will be detected. The solutions that we will deploy to the problem of rules and policy accountability are very closely related to the larger problem of assessing the reliability and provenance of information used in the intelligence analysis process. While we do not propose to solve the provenance problem, our work will have long synergy with and contribute to those efforts.

¹ Note early proposals for a 'privacy appliance.' Lunt, Teresa (2003). "Protecting privacy in terrorist tracking applications." Presentation to the Department of Defense Technology and Privacy Advisory Committee, September 29, 2003. www.sainc.com/tapac/library/Sept29/LuntPresentation.pdf

Section 3: Technical Approach

I. End-to-End Semantic Accountability

The goal of this project is to design Web-scale information architecture that is able to assess both data access requests and information usage events at the appropriate time, according to the rules relevant to that information. Web-scale operation is an operational requirement, first because of the breadth of information sharing, both in the public and the private sector; and second, because of the heterogeneous semantics of the data. In order to build a system that can operate at Web-scale, we look to the fundamental principle of Internet and Web design: the end-to-end principle [SRC84]. In order to provide meaningful accountability that relates actual uses of data to rules governing information, we will use the expressive power of Semantic Web techniques. End-to-end Semantic Accountability, then, will provide accountability across a broad spectrum of activities, conducted over highly diverse set of information.

The need for end-to-end semantic accountability is illustrated by considering the unique policy and trustworthiness requirements of an interagency information-sharing environment that links data from classified intelligence sources, sensitive civilian law enforcement records, Federal and State agency records, and commercial data sources, along with publicly available data such as that residing on the World Wide Web.

A. Policy Requirements and Threat Models

Increased integration of data from an ever-growing number of sources is the basic dynamic that drives the need for semantic accountability. Without delving into the classified details of national security investigation and analysis, we know that the following classes of data are collected and analyzed. Thus, it is a requirement to enforce rules across information that includes:

- Intelligence data: signals intelligence (intercepts and traffic data), raw human intelligence reports, analyst reports, watch lists;
- Civilian agency and criminal law enforcement data: investigation files, DHS/TSA Passenger Name Records, arrest warrants, electronic surveillance data (telephone wiretaps, email, Internet usage logs), drivers license records;
- Commercial databases: credit reports, last known address records, employment history;
- Public data: real property records, Web data.

Each of these classes of data carries with it a variety of rules, many of which are described with reference to the semantic content of the data, not its source, location or owner.

Distinguished observers of and participants in the debate over how to promote information sharing agree that technical infrastructure will play a pivotal role in enabling the operation of an efficient and trustworthy information sharing environment. A recent report from the Markle Foundation's Task Force on National Security in the Information Age found:

"Developing a trusted information sharing environment requires designing and implementing a technical infrastructure, and employing technologies with features that support policy goals. Overall architecture, technical applications, and features must provide opportunities for rules-based operation of systems in which the technology itself serves an important function to both enable appropriate, and constrain inappropriate, information sharing. In this sense, technology is an important enabler of trust..."²

The sentiment is echoed by The Honorable John Grimes, Chief Information Officer, Department Of Defense at the recent ODNI conference on information sharing. Speaking on the culture of information sharing and secret-keeping inside the various intelligence communities, he noted:

² "Mobilizing Information to Prevent Terrorism," Markle Foundation, July 2006. pp. 14-15 (http://www.markle.org/downloadable_assets/2006_nstf_report3.pdf)

"Unfortunately, we have individuals out there in the communities on both sides [ie. intelligence and defense] or all around that think they own the information. They do not own the information. They're just stewards of that information of the United States government and should be used as appropriately as possible."³

Therefore, the key threats against which we will design our end-to-end accountability infrastructure center around the dual risk that information may be shared too widely and misused, together with the mirror image threat that uncertainty about whether particular transfers or uses of information are permitted may *inhibit* what would be both legal and beneficial sharing and analysis. From these and many other assessments of the current status of information sharing and what is needed to proceed, the systems require the following capabilities:

- **Assure proper internal use:** In order to protect sources and methods, queries of all data sources should be able to enforce rules of the form: *inferences for the purpose of domestic air travel watch list creation are not allowed to combine Sigint from source A with humint from source B*. Note that while the rule includes traditional notions of data source limitations, full evaluation requires semantic characterization of the purpose or class of inference being conducted.
- **Reduce uncertainty in information sharing:** As a matter of national policy and a response to the experience of 9/11, information systems must inspire enough confidence to allow information sharing where legally authorized and, at the same time, discourage improper sharing that goes beyond the bounds of what is legal. This balance is vital to maintain. Just as systems ought to flag improper uses or transfers, they ought equally to engender confidence in the Intelligence Community that if a transfer or use is not flagged as improper, the intelligence worker should be able to proceed with the analytic mission.
- **Assess reliability of inferences from multiple data sources:** Aggregating data from a number of different sources also makes the assessment of the reliability of inferences very difficult. In the process of aggregation, individual reliability measures may be lost or hard to track by purely manual means. Semantic accountability is needed to provide rules such as: *Source A should be trusted for information regarding Islamic fundamentalist groups but not for assessments of nuclear WMD capability*. Human judgment will still be required to make final reliability determinations, but tools that provide provenance throughout a multistage, multisource process will help a user identify potential sources of uncertainty as well as to recognize multiple reports deriving from the same source.
- **Enable audit for legal compliance:** Just as tracking provenance in large-scale aggregation of data is complex, so too is managing and assessing compliance with legal rules. The interaction of the variety of legal rules that come along with a diversity of data sources thus requires machine assistance to help a user assure compliance with, and accountability to, all of the applicable rules.

These capabilities thus require accountability in which the system is able to evaluate information-usage and compliance-rules with respect to the meaning of the information in the system, as opposed to solely the source in which that information resides; i.e. *semantic accountability*.

B. Architectural requirements for Semantic Accountability

The three critical architectural requirements for semantically-accountable information flow are: first, a semantic framework within which to express both data usage and access control rules with reference to the content of the information over which accountability is sought; second, techniques to reason about

³ The Honorable John Grimes, Chief Information Officer, Department Of Defense, remarks at The DNI's Information Sharing Conference & Technology Exposition. Intelink and Beyond: Dare to Share (21 August 2006)
http://www.dni.gov/speeches/20060821_1_speech.pdf

rules compliance; and, third, comprehensive technological support for verification and audit of the associations between uses of information and the rules that allegedly authorized them.

Traditional security techniques fail to meet these requirements. With respect to creating a link between rules and the semantics of underlying data, role-based access control systems are entirely inadequate, and even rule-based access control systems do not have the ability to refer in a flexible manner to the semantics of information under control. With respect to the latter, no system that controls access to information *a priori* can either enforce or audit against rules that are only determined to be applicable based on uses that arise after access to information has already been granted. Hence, the requirement for late binding of rules must be met with a novel architecture. Work currently underway (See Section 3A) at MIT, Yale, Rensselaer Polytechnic Institute and the University of Maryland⁴ suggests that we can meet this requirement using Semantic Web, ontology, and rule-based reasoning techniques.

C. Existing security approaches will not provide end-to-end accountability

Sensitive data abound in today's networked world. By "sensitive data," we mean electronic data records that, if used improperly, can harm data subjects, data owners, data users, or other stakeholders. The profusion of sensitive data in a variety of public and private network environments and its increasingly pervasive role in everyday life are extremely important developments with wide-ranging social and legal consequences. Very robust technological trends (e.g., the plummeting cost of mass storage and the build-out of broadband networks) ensure that its potential misuse will continue to be a central concern for people and organizations.

Traditional security research (including but not limited to cryptologic research) takes a *preventive* approach to this very general technical and cultural challenge. Most existing security technologies strive to *limit the transmission* of sensitive data. Encryption, access control, and privacy-preserving, distributed function evaluation are well studied techniques that exemplify this approach. Note that effective use of *any* of them requires detailed understanding of the data, the users, the proposed uses, and the relevant laws and policies.

It is our thesis that there are fundamental reasons that a purely preventive approach cannot work. We briefly explain two of them here.

First, many sensitive data items are simply "facts" about people and organizations that are (1) legitimately used in a very large number of contexts and (2) fairly easily acquired by someone determined to do so. Although it may be reasonable to expect that an x-ray of one's shoulder remain in the records system of the radiology lab that created it and be seen only by medical-service providers who need to see it, it is not reasonable to expect that one's name and address will remain confined to a small number of isolated systems. Even if 99% of the organizations who acquire this succinctly representable, sensitive information use it appropriately, the negligent 1% could cause a deluge of unwanted communication; furthermore, a determined adversary can acquire this information with a modest amount of targeted effort.

Second, there is a general argument in favor of regulating *use* of sensitive information instead of inhibiting its dissemination. Determining whether a particular use is appropriate may require a great deal of information that is only available in context, i.e., at the time and place of the proposed use. While there are many differences between privacy and copyright laws, copyright provides a good example of the advantage of regulating use rather than access or collection. Draconian DRM systems that prevent access to such works can prohibit fair use, which is legal under copyright law, in an attempt to inhibit infringement, which is illegal. At the same time, determining whether a proposed use is allowed by the fair-use doctrine requires a great deal of contextual knowledge. Reasoning by analogy with the End-to-

⁴ Professor James Hendler, PI on the ongoing Policy Aware Web project and proposed co-PI here, will be moving from the University of Maryland to Rensselaer Polytechnic Institute in early 2007, hence his participation in this project is proposed through RPI not UMD.

End Arguments of networking, one is led to the general pattern of handling sensitive data in a manner that makes it available to all who may have a legitimate use but at the same time requires that potential users prove that they have a right to use it before doing so or at least be held accountable for what they have used it for.

At first glance, it may seem as though the need for end-to-end accountability applies to the open Internet but not to “closed” or “semi-closed” networks such as those used by the Intelligence Community. Unfortunately, the accountable-information-flow problem within the US Government’s Intelligence Community is not inherently easier than the accountable-information-flow problem in the open Internet. Any community that is comprised of multiple agencies with related (but not identical) missions, related (but not identical) data-use policies, and numerous independently administered computer systems will have trouble fulfilling its mission if it relies solely on preventive information-security technology to control (i.e., inhibit) the flow of mission-critical information. Moreover, it is even more important within the Intelligence community than it is in many open-Internet applications *not* to prevent the flow of potentially relevant information in situations in which one is unable to construct authorization proofs before the fact, and it is similarly more important to be able to construct proofs for accountability purposes after the fact.

Some of the research challenges that we will address in this project have been touched upon before in the security literature. For example, the *trust-management* approach to authorization enables the use of complex policies that are authored in a distributed fashion by multiple parties to a transaction, not all of whom know each other before the fact [BFIK, BFL]. Li and collaborators have combined the trust-management approach with traditional role-based access control [LMW] and have made some progress on *discovering* policy elements and credentials in situations in which a final decision maker cannot communicate with the sources of these policy elements and credentials [LWM]. Interestingly, REFEREE, an early example of a trust-management system for web applications [CFLRS], foresaw the need for compliance-checking procedures that do not always make a “yes/no” decision but sometimes return a result of “don’t know,” together with the potentially relevant evidence and reasoning that survives the attempt to make a decision. To date, however, no single framework or system combines all of the features necessary for end-to-end semantic accountability; we will re-use relevant features of existing trust-management systems whenever possible, along with relevant products of the PAW, TAMI, and PORTIA projects (See Section 3A).

II. Research challenges in accountable information architectures

End-to-end Semantic Accountability requires the ability to track information as it moves between applications in the distributed information environment that is today’s Web-based computing, for that information to be annotated in a form that lets us reason about its use, and for cryptographic means to be used both to verify identities in the distributed system and to provide secure auditability and proof verification. The PIs in this project have developed techniques in three different large past projects, and the goal of this work is to integrate these results into a unified whole that provides the end-to-end capability in a scalable and distributed way. The work will be based on:

- Semantic-Web Embedded Accountability: The “Policy Aware Project” (PAW; NSF ITR, PIs:Hendler, Weitzner, Berners-Lee), has been developing techniques for using rule-based reasoning to provide access control directly in the HTTP protocol using Semantic Web rules and reasoning. [WHBC05]
- Transparency of information usage and accountability of rule compliance: Developed in the Transparent, Accountable Data Mining Initiative (TAMI; NSF, PIs: Weitzner, Abelson, Berners-Lee, Fikes, Sussman) this work focuses on how scalable rule-based techniques, including those developed in the PAW project, can be applied to tracking the use of information and providing accountability in the Web environment. [WABH06]
- ID protection, verifiable provenance, and policy expression: The “Privacy, Obligations and Rights in Technologies of Information Assessment” (PORTIA; NSF ITR, PIs: Boneh, Feigenbaum, et al.) has been exploring state-of-the-art cryptography-based and logic-based methods for client-side identity protection [CLTBM04] and “contextual-integrity” based expression and application of sensitive-data policies and laws [BDMN06]. Recent collaboration by the PORTIA and TAMI

projects [FW06] has been exploring the use of PORTIA technologies for fortifying policy-use associations in decentralized information environments.

The success of our pairwise interactions, between PAW and TAMI and between TAMI and PORTIA, have shown us that it is possible to integrate these technologies into a larger-scale whole that can provide end-to-end accountability for information use. In the remainder of this section we outline these independent technologies. In Section 3A, we describe some of the specific outcomes of these projects, and here outline the research plan for integrating all these capabilities and proving they provide the capabilities described above.

A. Infrastructure-level support for accountability

Today's Internet architecture does not provide adequate support for the semantic, policy-based accountability required for large scale information sharing applications. First, Internet architecture is based on network addresses, not names. The binding of names to addresses is neither secure nor verifiable, and the same is true of the binding of transmitted data objects to addresses. Consequently, high-level authentication, authorization, and accountability mechanisms that are based on names (as they inevitably will be, because application semantics are expressed in terms of names, not network addresses) can be completely subverted by network-level attacks: denial of service, IP spoofing, DNS spoofing, etc.

The Policy-Aware Web (PAW) project has taken some first steps toward augmenting today's web with some "accountability infrastructure." [WHBC05] A remaining challenge, which we hope to address in this project, is to fortify PAW's associations of web resources and the policies that govern them with cryptographically supported proofs of ownership, integrity, timeliness, and other relevant properties.

The second failing of current Internet security techniques is the inability to prove compliance with *any* given policy. We will investigate how one can create network resources with universally understood, secure, persistent, and verifiable properties. What are the minimal sets of network resources upon which one can build proof systems for authorized use of sensitive data in a broad range of applications? Enabling accountability to rules that describe permissible and impermissible uses of information requires identification of data and responsible parties at a higher level of abstraction than can be provided by the network layer. Even if we could securely bind names to addresses and securely bind classes of agents to names, we would still need to discover (a consistent and interpretable representation of) the relationships among agents and classes of data and the rights and responsibilities of the relevant agents.

Here, too, results from the Policy Aware Web project suggest directions that we propose to pursue further in this project. PAW associates Web-based resources with the policies that govern their use by adding an extra step to HTTP. [KKHWB05] When user U issues an HTTP GET request for resource R_1 whose URI is A_1 , the controller of R_1 first sends to U the policy P_1 that governs access to R_1 . The user must then assemble a proof that, according to P_1 , he is authorized to access R_1 . He sends the proof to the controller, who verifies it; if this verification step succeeds, then U is allowed to download R_1 from A_1 .

Although this is a good first step toward end-to-end semantic accountability, it is not sufficient. Remaining issues that we will address in this project include but are not limited to:

- **Security of policy associations:** Protocol-level attacks (e.g., session stealing or man-in-the-middle attacks) may leave the PAW framework vulnerable to U's receiving an erroneous policy $Q_1 \neq P_1$ from an attacker (and thus failing the verification step, because he has assembled and sent to the controller a proof that he is authorized by Q_1 , not by P_1) or to his receiving resource $S_1 \neq R_1$ from an attacker after passing the verification step.
- **Multiple sources of authority:** Note that the basic PAW approach assumes that P_1 and R_1 are controlled by the same entity, say C_1 . In some scenarios, however, multiple entities must grant permission before U can access R_1 . Moreover, after U downloads R_1 and uses it in a computation, he may need to access another resource R_2 controlled by C_2 and governed by P_2 . Both the results of U's earlier computation and the proof that he was authorized under P_1 to

download the necessary resource may be needed to satisfy P_2 . How should we deal with the cases in which C_2 does not have ready access to P_1 ?

- **Multi-phase computation and authorization:** As noted in the above discussion of multiple sources of authority, an action for which U will be held accountable may actually be a sequence of actions, involving multiple resources, multiple controlling authorities, and multiple policies. How can the entire *chain* of evidence, together with all relevant temporal and causal relationships, be preserved in a secure fashion?

In this part of our project, we will leverage the large body of work that the cryptologic-research community has done on relevant techniques, e.g., secure logging, secure time-stamping, and multi-round authenticated sessions. One of our goals is to enable a user (when he is called upon to justify after the fact a particular use of sensitive data) to retrieve, among other things, signed and dated records of all actions he took, all policy-governed resources that he used in each action, the relevant policies and proofs of compliance that he provided, and the proofs that he received and verified that the provenance of these policies and the resources they govern were in order.

In principle, all of these aims can be achieved with known cryptographic-protocol techniques. In practice, however, no one has ever sewn all of the secure components together to form an up-and-running, end-to-end secure system for semantic accountability, and doing so will be nontrivial.

At this point, it is crucial to make clear what we will *not* attempt to do in this project. Simply put, we do not plan to develop solutions to extremely low-level system-security problems or to extremely high-level organizational-security problems. Unconditionally secure transaction logging, for example, may require the development of custom-designed secure hardware or operating systems, and these would be beyond the scope of our project. We will point out the aspects of our solution in which commercially available secure hardware or operating systems would be useful and where something new would be needed, but we will not design new hardware- or OS-level solutions. Similarly, we will assume throughout the project that each constituent intelligence agency has a basically sound leadership team and IT staff; thus, standard techniques to thwart small-scale “insider attacks” should suffice. Agency-wide corruption or incompetence could undermine our solutions, e.g., by poisoning the PKI or the policy database. It is our belief that this technical scope is consistent with NICECAP BAA 06-11 and with the term “semantic accountability.”

B. Scalable Reasoning in Accountability Appliances

End-to-end accountability will require the widespread deployment of 'accountability appliances' across the network. Described in more detail in Section 3.III, these appliances will be deployed at endpoints in the network, close to users on the one hand, and data sources, on the other. The design and interaction of these accountability-assessing components poses substantial research challenges regarding the choice of reasoning mode, degree of policy expressivity allowed, and the binding between these small appliances and the overlaying information analysis applications. From the broad challenge of designing a scalable approach to accountability reasoning, we propose to address three particular research questions:

1. Can we design accountability appliances that can be distributed throughout a large information network while still enabling overall reasoning about policy compliance?
2. Can we apply truth maintenance system (TMS) techniques to preserve a running record of the compliance status of information usage?
3. What kinds of constraints will be necessary when changes in rules and facts occur after information has already been used?

We discuss each of the research questions here based on our experience in the Transparent Accountable Data Mining Initiative (TAMI).

1. Distributed reasoning modules

Accountability appliances must be able to reason about the restrictions on the use of information, and how those restrictions must be combined when corresponding pieces of information are combined. If the language of restrictions is unlimited in expressive power then the problem of tracking and combining restrictions is computationally infeasible. The quantity of information used in an organization scales, perhaps nonlinearly, with the size of the organization. Therefore, the computational and communication resources required for tracking that information and manipulating the restrictions on its use must scale appropriately. Our design hypothesis is that we can obtain this kind of scaling by a distributed architecture. We believe we will succeed with this approach because restrictions on the acceptable uses of information are sufficiently circumscribed that there are techniques for controlling the computational complexity of the manipulations. However, this is still an open research question.

So accountability appliances will be distributed throughout the institutional infrastructure. Some of these appliances will operate incrementally, reasoning about data as it accumulates. Others will be used to reason about long connected chains of data after the fact. However, simple distribution of the appliances isn't enough. We must constrain the individual appliances so that each operates mostly independently of the others, otherwise the interactions between the appliances will cancel out the advantages of distribution. Fortunately the computation of usage restrictions is mostly local in nature, which reduces inter-appliance communication. Restricting the inputs and outputs of the appliance to write-once (monotonic) storage eliminates inter-appliance locking and allows wide-spread caching of non-local data, which further reduces communication.

2. Truth Maintenance Systems to preservation of intermediate policy conclusions

Through the network of distributed accountability appliances, conclusions must be preserved at least as long as the data from which they are derived. In our initial work on the TAMI project, we have begun to apply Truth Maintenance System (TMS) techniques [StSu76] to the problem of assessing policy accountability in the case of specific national security/law enforcement data sharing scenarios based on actual Privacy Act notices.[WABH06] It must be possible to modify these conclusions in the presence of new data, or improved reasoning engines. However, for accountability, the data store should be monotonic: no data is ever deleted; modifications must be made by annotation of existing data items.

Appliances must be able to cooperate across institutional boundaries. Agency policies place restrictions on data flow across their boundaries, and the appliances must conform to these policies while doing their jobs. For example, an appliance running inside the CIA must not reveal classified information when cooperating with an external agency. Additionally, the appliances and their conclusions must be able to adapt to policy changes.

3. Constrained reasoning

Data always has usage restrictions. Data that is acquired from outside the system often comes with explicit restrictions, for example by statute or contract. Additional restrictions may be accrued during transfer of data from one place to another within the system. And combinations of data have restrictions that are functions of the incoming restrictions and the means of combining. A consequence of this structure is that the restrictions on a data item are computable from the restrictions imposed by the original data sources and by the ways that the data is combined or transferred.

However, the rules for usage restrictions sometimes change, which means that previously computed restrictions must be updated to reflect the new rules. For example, a court order may require data to be

used in previously disallowed ways. Or enhancements to privacy law may add further restrictions to the usage of personal information. These kinds of changes may require extensive computation.

Fortunately there are some constraints that limit the scope of such changes. For example, a court order refers to a particular case, affecting only the usage of data related to that case, which is usually a tiny slice of a large data store. Statutory changes are more pervasive, but even so it may be possible to use explicit dependency structures to quickly locate the areas that are affected by such a change. We will address the simple changes, such as court orders, in this project, leaving more complex changes for future work.

III. Building and Deploying Accountability Appliances

A. General Architecture for End-to-End Semantic Accountability

We propose to design, build and test a general-purpose accountability appliance that will be able to scale up to operation for arbitrarily complex information spaces. The functions of this appliance will include:

- o **Policy-aware access control:** the accountability appliance will mediate requests for data coming from users. Access decisions will be made by evaluating proof of access rights generated by the user's accountability appliance using the underlying semantics of the data being requested in conjunction with applicable rules from relevant institutional data owners and relevant external rule sources (legal requirements, etc.).
- o **Provenance-aware reliability assessment:** In processing queries, the accountability appliance will assess the reliability of the results based on the relationship between the data sources and rules describing the reliability of those data sources in the context of the particular queries made and ultimate inferences drawn. Users will have access to general provenance/reliability rules, be able to build their own rulesets, and discover and share rules from other users.
- o **Policy-and provenance-aware 'what-if' tools:** In building complex chains of inferences, analysts and intelligence users will want to assess whether their uses of data comply with relevant usage and information-sharing policies. Some of these policies may be put in place by particular data owners to protect sources and methods; other policies will result from the legal requirements to protect individual privacy rights as established by legislative and executive branch authorities. As users proceed with an investigation or analytic task, they will be able to request an evaluation of their compliance with the relevant rules.
- o **Policy-aware audit and accountability:** Given the dynamic nature of national security investigations, it may not be possible to assess rule compliance until the investigation is complete. Users will be able to exploit the late-binding of information rules to the results of an investigation, both for the purpose of establishing legal compliance, and as a final test of the reliability of their results.

We illustrate the design of the system that we plan to build over the a three to five year period in Figure 1.

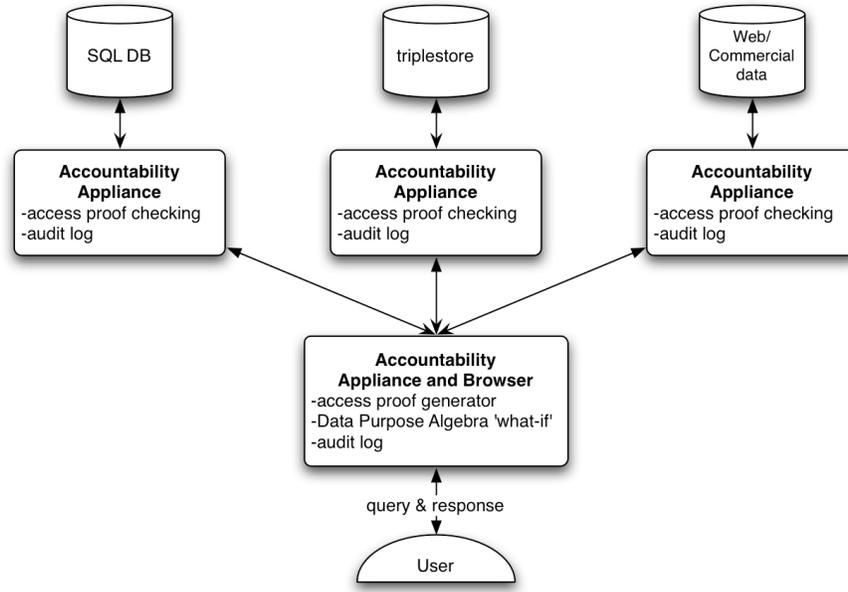


Figure 1: Architecture for End-to-End Semantic Accountability

B. Phase I Deliverables – A working prototype of the End-to-End Accountability Environment

In Phase I of this project, we propose to build a working prototype of each major component of the End-to-End Semantic Accountability architecture that we have described. Given belief that the architecture of the World Wide Web (and the Semantic Web in particular) is the appropriate design for the needs of this project, the specifications and software components we build will all be implemented as Web infrastructure components, including Web server and client-side proxy modules. This work is logically divided into three deliverables:

1. Specification of the Data Purpose Algebra, a language with which we can describe permissible uses of sensitive information.
2. Client-side http proxy, constructing access proofs and logging transactions in a secure manner and Apache Web server module checking and acting on access requests (using a proof checker), providing secure logging of all information access and usage transactions and depositing these transaction records into a TMS.
3. General accountability browser, providing users the ability to evaluate whether or not a given use of a collection of information complies with the relevant rules and policies.

Together, these deliverables will comprise what we have called Accountability Appliances that can be installed in various places across the Web. We describe our current plans for each deliverable in more detail here.

1. Data Purpose Algebra

Data is often encumbered by restrictions on the ways it may be used. These encumbrances may be determined by statute, by contract, by custom, or by common decency. Some of these restrictions are intended to control the diffusion of the data, while others are intended to delimit the consequences of actions predicated on that data. The allowable uses of data may be further restricted by the sender: "I am telling you this information in confidence. You may not use it to compete with me, and you may not give it to any of my competitors." Data may also be restricted by the receiver: "I don't want to know anything about this that I may not tell my wife."

Although the details may be quite involved, as data is passed from one individual or organization to another the restrictions on the uses to which it may be put are changed in ways that can often be formulated as algebraic expressions. These expressions describe how the restrictions on the use of a particular data item may be computed from the history of its transmission: the encumbrances that are added or deleted at each step. A formalization of this process is a Data-Purpose Algebra description of the process. One pervasive assumption behind our formalization is that a data item, annotated with its provenance, may be restricted, but this restriction is not on the content of the data item. For example, a law-enforcement official may not act on improperly obtained evidence, but if the same information was redundantly obtained through lawful channels the official may act. Of course, there are other real-world circumstances where this assumption is invalid. For example, consider the fact that Joe ate ice cream at 3:12 PM on 13 August 2006. Now suppose that in the playground Anne told me that Mary told her that Mary observed that Joe ate ice cream at 3:12 PM on 13 August 2006. Don't tell his mother! We see that this item has restricted distribution. But if also Jim told me that Joe ate ice cream at 3:12 PM on 13 August 2006 our formalization would allow me to tell Joe's mother that he ate ice cream before dinner. However, I would feel inhibited, as a matter of courtesy, by the fact that Anne told me not to pass this information along.

In the illustration that follows we consider a simplified formulation of the rules for data passed among government agencies and officials, specified by the Systems of Records Notices associated with Systems of Records, as defined by the Privacy Act. Each data item i has, in addition to its content $q = Q(i)$, a set of purposes $p = P(i)$ for which it can be used. An item is constructed from its content and its purposes $i = I(q, p)$. Let s be a data source, for example, a System Of Records (a SOR). A system of records may be controlled by an organization $c = C(s)$. Also associated with the system of records may be a System Of Records Notice (a SORN) $n = N(s)$, which gives information about the permissible uses of the SOR.

If there is a SORN, it specifies a set of source-use entries $e = E(n)$ for data extracted from that SOR. Each entry $e \in e$ specifies a set of possible recipient organizations $O(e)$ and the set of authorized purposes $U(e)$ for which the specified recipient organizations may use data from the source. Any particular recipient r_1 may be a sub-organization of a possible recipient organization r_2 specified in a SORN. This relation is notated $r_1 < r_2$. Thus, the set of applicable source-use entries $A(s, r)$ for transfer of data from a source s to a recipient r is just the set of those entries for which the recipient is a sub-organization of an organization specified as a recipient organization of an entry in the SORN:

$$A(s, r) = \{e \in E(N(s)) \mid \exists o (o \in O(e)) \wedge (r < o)\}$$

The restriction on authorized purposes of a transfer from a source to a recipient is that the purposes must be authorized by any of the entries that contain the recipient organization.

$$R(s, r) = \bigcup_{e \in A(s, r)} U(e)$$

The authorized purposes $Z(s, r, i)$ to which a recipient r may put a data item i extracted from a source s is then restricted to be those purposes particular to that data item that are also allowed by one of the purposes in the authorized routine purposes obtained from the SORN:

$$Z(s, r, i) = P(i) \cap R(s, r)$$

So $Z(s, r, i)$ is the set of purposes of the new item held by the recipient r with the content of the old item i held by the source s . The transfer of a set of items i from a source s to a recipient r is a new set of items $T(s, r, i)$:

$$T(s, r, i) = \{I(Q(i), Z(s, r, i)) \mid i \in i\}$$

When formalized in this way, computations described by these algebraic descriptions are directly represented as purely functional computer programs. There is no complex translation required.

The example shown above is incomplete in that it does not cover all of the requirements of the Privacy Act: it does not cover restrictions on the actual collection of data for a SOR imposed by the SORN, and it does not distinguish between the categories of data defined by the SORN. These are easy extensions to the formal descriptions shown above.

But there are harder problems. We have not begun to consider the informal and implicit restrictions on the use of data required by cultural considerations, such as courtesy. However, we must confront the problem of being able to formally describe such currently informal notions to ensure that we can make a system that is sufficiently general to cover real-world situations. Another problem is revealed by the situation where an entity is allowed to discuss the consequences of a secret it knows with any other entity that already knows that secret. Similarly, it is possible that an entity may hold a secret that it is only allowed to divulge if the reality is that the information is generally available through other channels.

In general, the algebraic approach is well suited to modeling the allowable uses of information when the restrictions on that use are determined by the path by which the information is obtained, but it is not so good at dealing with restrictions that are time dependent or inherent in the content of the information, independent of the path. We will have to design means of modeling information with time-dependent and content-dependent restrictions.

2. Client-side proof generating proxy and server side accountability module

In section III, we described the use of Semantic Web technologies to couple policies with resources on the Web. A key component in doing this, and one which will be used as the basis for integrating information resources into the Semantic Accountability Appliance, is a Semantic Web proxy which is able to find policies when encountering 401 errors and, in PAW, to generate proofs and to send these to the server which governs the use of the resource. [Figure 2] outlines the use of this proxy/server component in PAW.

For PAW, the server uses a particular permissions language (REIN) to describe policies and a rule-based reasoner, called CWM, which is able to both generate and check proofs. The proxy is designed, however, in such a way as to make it possible for other technologies to be used instead of these. In the current work, we propose to extend this proxy to work with the data-purpose algebra and with cryptographically supported proof systems. As such, it forms an important component for providing an interoperability mechanism, embedded in the Web architecture, for associating resources, policies and keys as needed in the accountability appliance.

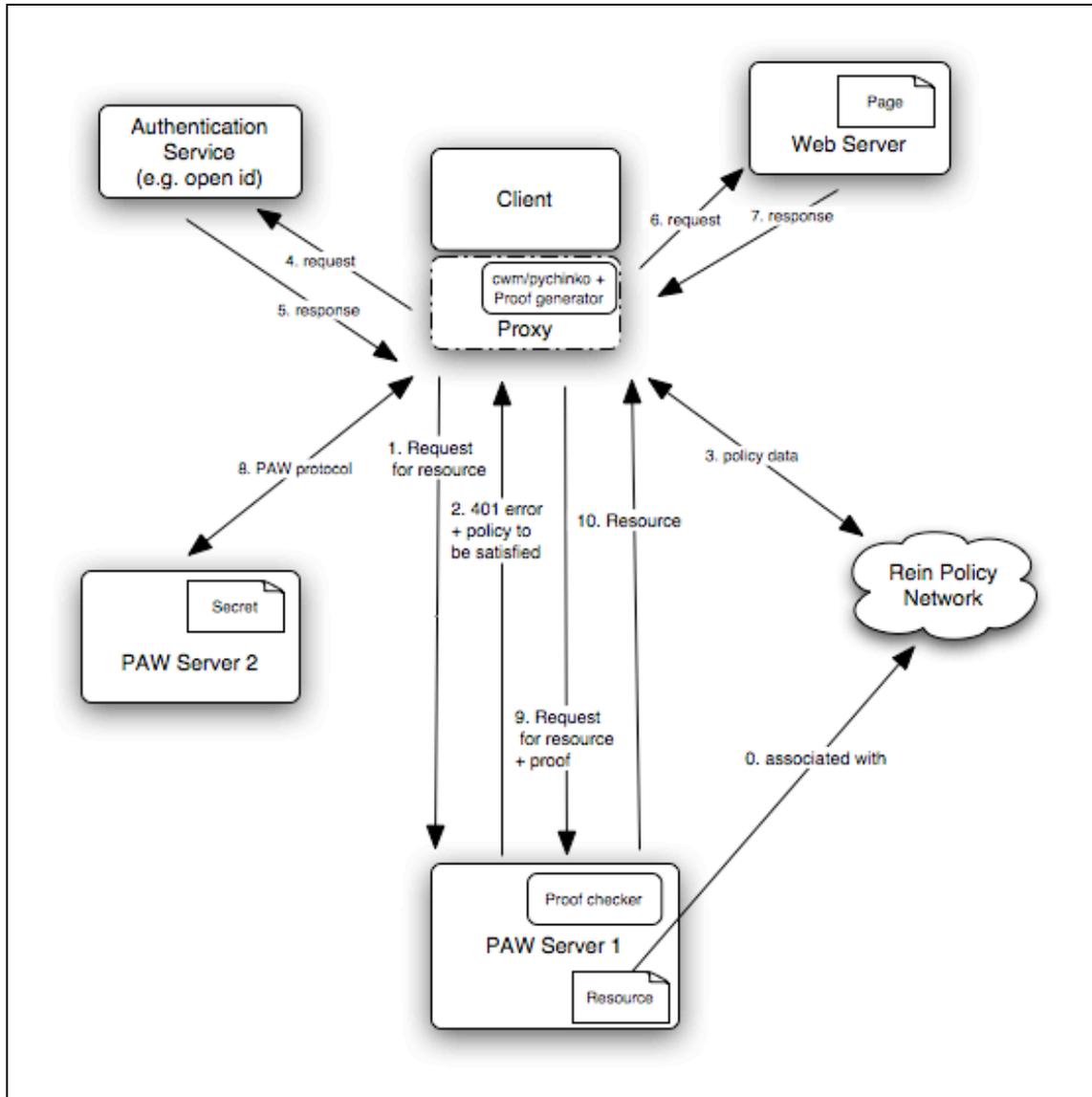


Figure 2

Both the client and server-side components will use scalable distributed reasoning module adequate to handle accountability computations for data-use restrictions. This system will be designed with standardized component interfaces, allowing individual components to be replaced or combined with others. For example, reasoning engines should be composable and interchangeable. A consequence of this is that reasoning must be incremental and monotonic.

Appliances will be distributed both in space and in time. Distribution in space means they must be able to cooperate with one another while reasoning about common data. Distribution in time means that future appliances must be able to use the conclusions reached by those in the past, and that past appliances should facilitate the actions of those in the future.

Individual appliances must be able to trust the conclusions generated by other appliances. End users of the system must be able to trust the conclusions generated by the system as a whole. Consequently there need to be mechanisms to allow the outputs of the appliances to be evaluated for trustworthiness. This requires that all results must be cryptographically signed. These results must also carry appropriate provenance information sufficient to give confidence in the validity of the results. However, there are

situations where the provenance information must not be released with the results. We believe that cryptographic-protocol techniques, e.g., zero-knowledge proof systems, may be useful in this context.

3. General accountability browser

We propose to implement a specialized Semantic Web browser (an Accountability Browser) that can render accountability reasoning, provide useful cross references, and allow annotation, editing, and publishing of the data and reasoning presented. A chain of reasoning will rarely be useful in isolation, however. Investigations into complex legal matters will ultimately require the checking of references and the verification of facts. The tools developed at MIT for interacting with the Semantic Web may prove to be a useful starting point. Allowing observers to browse the system will allow observers to cross-check assertions made by the reasoner against commonsense legal rules. To the extent that government databases can easily be made accessible to the judges making decisions about the evidence, we will design an interface to allow for judges to investigate them using a simple interface.

C. Phase II – End-to-End Accountable Systems at Large Scale

Phase 2, if funded, would further this demonstration in several areas of scalability: Tools for the easy addition of new data sources would be developed, including semi-structured and unstructured as well as structured databases. The rule-based reasoning capabilities used in the accountability assignment will be optimized for both efficiency of operation and compatibility with emerging Web standards. The number and complexity of policies will be extended and proven to be sufficient for the representation of critical information-sharing policies used in the government operations.

D. Phase III – From Prototype to Operational Deployment

Phase 3 work, and the evaluation thereof, will focus on the transition of this technology for use in meeting the information accountability needs of the National Intelligence Community. The University collaborators performing the research described in this proposal would require teaming with an IC contractor able to harden and deploy the techniques described herein.

Section 3A: Results of related prior and current government-funded R&D efforts

The work proposed here builds on our ongoing research in the Semantic Web and ontology fields, as well as ongoing work in the application of cryptographic techniques to policy management. Our current work explores the application of rules and policies in Web-scale decentralized information systems, as well as developing techniques for the protection of sensitive information in network environments. Through the Policy Aware Web project (NSF, PIs: Hendler, Berners-Lee, Weitzner, through September 2007) we have developed basic tools for controlling access to web-based information resources through use of proof-carrying authentication techniques. We propose to extend this work to rule sets that are of specific interest to the Intelligence community, and to evaluate the utility and scalability of these techniques in larger scale systems than is possible with our current project resources. In the Transparent Accountable Datamining Initiative (NSF, PIs: Weitzner, Abelson, Berners-Lee, Sussman, Fikes, through September 2008), we are investigating techniques for describing complex legal requirements in machine-readable rules. Reasoning and truth maintenance techniques, developed in the TAMI project, will provide an important foundation for semantic accountability in the systems described here. We will leverage work already done in the PORTIA project on techniques for protection of sensitive information.

Policy Aware Web (PI: Hendler, CoPI: Weitzner, Berners-Lee)

The Policy Aware Web Access Control project has recently completed a first demonstration of technology that can allow the sort of fine-grained, open, distributed and scaleable access control on the World Wide Web that is required by examples such as those above. Our approach provides for the publication of declarative access policies in a way that should allow for controlling the sharing of information without requiring the sorts of pre-agreement necessary in current systems. The project is aimed at putting

greater control over information into the hands of the information owners, be they scientists, government workers, parents of Girl Scouts or anyone else hoping to share information, but in a controlled manner.

The PAW project is part of an effort to explore Web-based technologies that enable *policy aware* infrastructure, that is, that allow people and organizations to more easily describe the conditions under which certain transactions can occur on the Web. The work makes use of Semantic Web technologies, which provide new languages and tools for the World Wide Web. Currently, most Semantic Web research is focused on data interoperability and ontological description of Web resources. The PAW research extends this work to explore the use of rule and proof languages in general, and for access control in particular. Thus, successful development and deployment of this technology will allow information resources on the World Wide Web to carry access policies providing for the continued wide dissemination of information without sacrificing individual privacy concerns. The PAW demonstration shows that it is possible to deploy rules in a distributed and open manner, and to produce and exchange proofs based on these rules in a scaleable way, providing a foundation for this new level of privacy control that can someday be deployed on the World Wide Web.

Transparent Accountable Data Mining Initiative (PI: Weitzner, Co-PIs: Abelson, Berners-Lee, Fikes, Sussman)

The goal of the TAMI project is to create the technical, legal and policy foundations for transparency and accountability of large-scale aggregation and inferencing across heterogeneous data sources. The TAMI Project is addressing the risks to privacy protection and the reliability of conclusions drawn from increasing ease of data aggregation from multiple sources by creating robust, scalable designs for adding increased transparency and accountability of the inferencing and aggregation process itself.

In our first year of work, we have pursued our work through a tightly coordinated team of investigators (PIs, co-PIs, senior personnel, students and a technical programmer) meeting on a weekly basis to explore concrete policy scenarios and build systems that address accountability and transparency needs in those scenarios. Through this collaborative work we have identified and begun to see results in three areas of inquiry:

a) What is the proper expressive framework for the classes of legal rules against which we seek to assess compliance? Here we have learned that the N3 language is likely to be useful, though higher order abstractions may be required for effective use. Some of the challenges we have encountered here are more social and institutional than scientific. In particular, the lack of a consistent naming convention for legal rules is a real practical barrier to machine-assisted reasoning.

b) At what point in a data mining process is it most effective to assess compliance? We began with a hypothesis that many of the query-limitation approaches often known as privacy preserving data mining' would not be sufficient to protect privacy in complex investigations. The early scenarios that we have developed already demonstrate that it is sometimes logically impossible to assess compliance with privacy rules until the very end of an inferencing chain. This requirement for late-binding of rules to data suggest that value of our transparency and accountability approach over those that seek complete privacy protection through a priori control over each individual query or data access event.

c) What are the most useful user interface metaphors to enable human users (judges, law enforcement investigators, etc.) to assess rule compliance?

PORTIA⁵ (PIs: Boneh, Feigenbaum, et. al.)

Since its inception in the fall of 2003, the PORTIA project, a large-ITR study of "Privacy, Obligations, and Rights in Technologies of Information Assessment," has been vigorously promoting the importance of

⁵ There are five universities, ten academic PIs, and numerous research partners involved in PORTIA, and so we give only the Yale personnel here. Please see <http://crypto.stanford.edu/portia> for a complete list.

supporting appropriate use of sensitive data rather than simply limiting its dissemination. PORTIA technical accomplishments to date that are most relevant to the work proposed here include but are not limited to: (1) development of the notion of contextual integrity, its instantiation in modal logic, and its application to actual US laws (including HIPAA and Graham-Leach-Bliley); (2) design and implementation of a novel database-query engine tailored for biosciences databases used both by clinicians and by multiple research teams at multiple universities; and (3) browser-based technology for identity protection. Note that all of these results and technologies aim squarely at the fundamental reality of our information environment: Networks are increasingly dominant precisely because they enable people and organizations to share information; hence, attempts to achieve “security” solely by hiding information will probably fail; many important applications are characterized by many parties with diverse and shifting interests, most pairs of which are neither completely allied nor completely adversarial, and most of whom have strong motivations both to share data for specific purposes and to discourage its misuse.

Section 4: Evaluation Approach

We propose to evaluate our work against these basic criteria:

1. Expressivity of our policy languages
2. Scalability
3. Resilience against identified threat models:
 - a. Ability to evade accountability without detection
 - b. Discouraging information sharing because of uncertainty about whether sharing is permissible in a given context.

Some of this evaluation will be done by our own project team, but we will also seek assistance from a candidate user institution to help with the evaluation of our policy languages.

Policy Language Expressivity

The lynchpin of the success of the end-to-end semantic accountability approach is to have a framework for describing policies that has both adequate expressive power and at the same time is amenable to efficient machine reasoning in support of accountability assessments. Hence, our evaluation of the expressive properties of our policy language will be done in cooperation with individuals from a user organization who has compliance responsibility. Thus we will be able to assess whether the language we design as part of the Data Purpose Algebra is adequate to the real work needs of at least one user in the Intelligence Community.

The framework we develop for expressing policies against which accountability will be measured must be able to express the types of rules and law currently in force in actual national security and criminal law enforcement investigations, and should have a reasonable likelihood of being able to express new rules that will be developed. Following the model we pioneered in the TAMI project, we will test the ability of our policy languages to encode actual United States laws and regulations. In subsequent phases we would also work with legal counsel from various agencies to further test the frameworks we develop against rules and laws that are relevant to various parts of the Intelligence Community.

Scalability Evaluation

The scalability properties we seek to satisfy, and will evaluate against, are twofold. First, we expect our accountability architecture to overlay on the existing World Wide Web. Therefore, we will deploy and test our tools on publicly-accessible Web sites and assess their performance running as part of production Web servers. Second, the reasoning tasks required to assess accountability must be able to operate efficiently across a number of independent accountability appliances. Again, deployment on a variety of publicly-accessible, dispersed Web sites will help us to evaluate computer power required to achieve desired accountability results.

Resilience against identified threat models

A central thesis of the project is that it should be possible, given secure maintenance of evidence chains, to assess accountability to rules that govern the actions recorded in those evidence chains. We recognize that there are a variety of entirely out-of-band infrastructure attacks possible to this system (e.g., steal a hard drive containing sensitive data or deploy brute-force computing power to break ciphers). We do not pretend to offer any new approaches to this type of problem, but we do argue that, if basic system perimeters are protected (through convention security means), our accountability mechanism will be able to spot information usage that deviates from stated rules. In order to test this proposition, we will develop formal models of the environment we are building and will attempt to prove that it is *infeasible* to hide improper uses of data in what appear to be otherwise permissible uses.

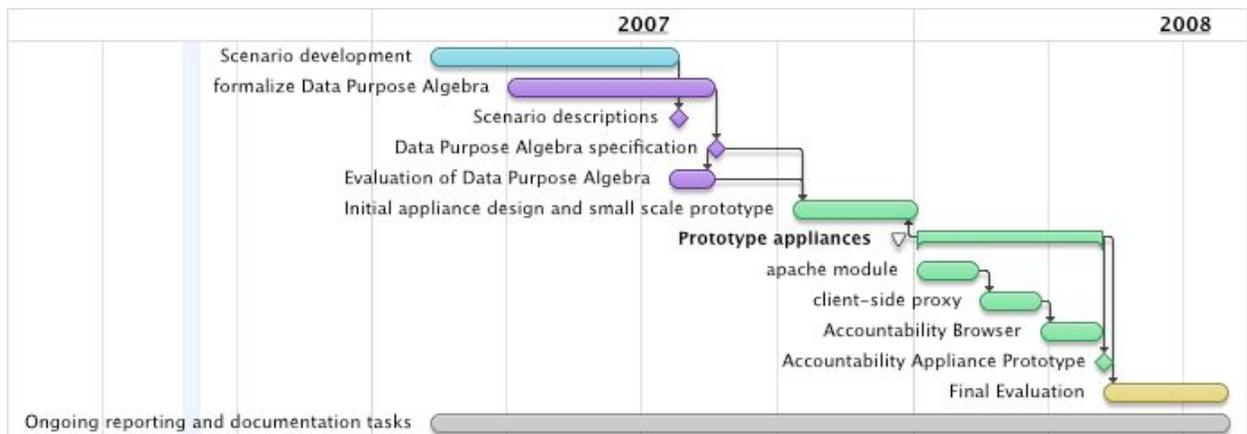
General Evaluation Approaches

In Phase I, all evaluations will be conducted by members of our own project team working with individuals we identify from one representative user community. In Phase II, if funded, these same evaluation criteria can be tested by independent users and implementers. To ease the evaluation of this work, all code developed under this project will be released open-source using licensing that allows unrestricted access for government evaluation and use.

Section 5: Schedule/Milestones

In the Gantt chart show here, we illustrate our plan for reaching the following major milestones of the project:

1. Scenarios describing information sharing activities that require accountability to specific rules and policies.
2. Data Purpose Algebra specification
3. Prototypes of Accountability Appliance components
4. Evaluation of our work (at various stages of the project).



Section 6: References

- [BDMN06] A. Barth, A. Datta, J. Mitchell, H. Nissenbaum, "Privacy and Contextual Integrity: Framework and Applications." Proceedings of the 27th IEEE Symposium on Security and Privacy, IEEE Computer Society, 2006.
- [BFIK] M. Blaze, J. Feigenbaum, J. Ioannidis, and A. Keromytis, "The KeyNote Trust-Management System, Version 2," *Internet RFC 2704*, September 1999. <http://ftp.isi.edu/in-notes/rfc2704.txt>.
- [BFL] M. Blaze, J. Feigenbaum, and J. Lacy, "Decentralized Trust Management," in *Proceedings of the 17th Symposium on Security and Privacy*, IEEE Computer Society Press, Los Alamitos, 1996, pp. 164 - 173.
- [CLTBM04] N. Chou, R. Ledesma, Y. Teraguchi, D. Boneh, J. Mitchell, "Client-side defense against web-based identity theft." 11th Annual Network and Distributed System Security Symposium (NDSS '04), San Diego, February, 2004.
- [CFLRS] Y.-H. Chu, J. Feigenbaum, B. LaMacchia, P. Resnick, and M. Strauss, "REFEREE: Trust Management for Web Applications," *World Wide Web Journal* 2 (1997), pp. 127 - 139. Reprinted from *Proceedings of the 6th International World Wide Web Conference*.
- [FW06] Feigenbaum and Weitzner (ed.), "Report on the 2006 TAMI/Portia Workshop on Privacy and Accountability." <http://dig.csail.mit.edu/2006/tami-portia-accountability-ws/summary> (August 2006)
- [KKHWP05] Towards a Policy Aware Web, in IWSC Semantic web and policy workshop, Vladimir Kolovski and Yarden Katz and James Hendler and Daniel Weitzner and Tim Berners-Lee., 2005.
- [LMW] Ninghui Li, John C. Mitchell, and William H. Winsborough. Design of a role-based trust-management framework. In *Proceedings of the 2002 IEEE Symposium on Security and Privacy*, pages 114–130. IEEE Computer Society Press, May 2002.
- [LWM] N. Li, W. Winsborough, and J. Mitchell, "Distributed Credential Chain Discovery in Trust Management," *Journal of Computer Security*, volume 11, number 1, pp. 35-86, February 2003.
- [SRC84] J. Saltzer, D. Reed, and D. Clark. End-to-end arguments in system design. *ACM Transactions on Computer Systems* 2, 4 (November 1984) pages 277-288
- [StSu76] R. Stallman, G. Sussman, "Forward Reasoning and Dependency-Directed Backtracking in a System for Computer-Aided Circuit Analysis," AIM-380: September 1976. <ftp://publications.ai.mit.edu/ai-publications/0-499/AIM-380.ps>
- [WABH06] Weitzner, Abelson, Berners-Lee, Hanson, Hendler, Kagal, McGuinness, Sussman, Waterman, Transparent Accountable Data Mining: New Strategies for Privacy Protection,; MIT CSAIL Technical Report MIT-CSAIL-TR-2006-007(27 January 2006).
- [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, 2005.

Appendix of Resumes

Hal Abelson

Education:

Princeton A.B. (summa cum laude) 1969
MIT Ph.D. (Mathematics) 1973

Professional Appointments:

1994–present MIT Class of 1922 Professor MIT
1991–present Full Professor of Computer Sci. and Eng. MIT
1982–1991 Associate Professor of Electrical Eng. and Computer Sci. MIT
1979–1982 Associate Professor, Dept. of EECS and Division for Study and Res. in Education MIT
1977–1979 Assistant Professor, Dept. of EECS and DSRE MIT
1974–1979 Lecturer, Dept. of Mathematics and DSRE MIT
1974–1979 Instructor, Dept. of Mathematics and DSRE MIT

Selected publications relevant to this proposal:

1. Structure and Interpretation of Computer Programs, Hal Abelson, Gerald Jay Sussman and Julie Sussman, MIT Press and McGraw-Hill, 1985, (published translations in French, Polish, Chinese, Japanese, Spanish, and German). Second Edition, 1996.
2. “The Risks of Key Recovery, Key Escrow, and Trusted Third-Party Encryption,” with Ross Anderson, Steven Bellovin, Josh Benaloh, Matt Blaze, Whitfield Diffie, John Gilmore, Peter Neumann, Ronald Rivest, Jeffrey Schiller, and Bruce Schneier, in World Wide Web Journal, vol. 2, no. 3, Summer 1997, O’Reilly & Associates, pp. 241–257.
3. “Intelligence in Scientific Computing,” Hal Abelson, M. Eisenberg, M. Halfant, J. Katzenelson, E. Sacks, G.J. Sussman, J. Wisdom, K. Yip, CACM, 32, no. 5, May 1989.

Selected other publications:

1. “The Supercomputer Toolkit: A general framework for special-purpose computing,” with A. Berlin, J. Katzenelson, W. McAllister, G. Rozas, G. J. Sussman, and Jack Wisdom, International Journal of High-Speed Electronics, vol. 3, no. 3, 1992, pp. 337–361.
2. Apple Logo, Byte Books, Peterborough, N.H., 1982 (alternate edition of Logo for the Apple II).
3. Turtle Geometry: The Computer as a Medium for Exploring Mathematics, with A. diSessa, MIT Press, 1981 (published translations in Spanish, Italian, and Polish).
4. “Amorphous Computing,” Harold Abelson, Don Allen, Daniel Coore, Chris Hanson, George Homsy, Thomas F. Knight Jr., Radhika Nagpal, Erik Rauch, Gerald Jay Sussman, and Ron Weiss, in Communications of the ACM, 43, 5, May 2000.
5. “Amorphous-computing techniques may lead to intelligent materials,” with Nancy Forbes, in Computers in Physics, vol. 12, no. 6, Nov/Dec 1998. Reprinted in Jour. Complexity, vol. 5, no. 3, January 2000.

Synergistic Activities:

Abelson's professional career centers around the use of computation as a framework for formulating knowledge in science and engineering, both to create better tools for science and engineering and to better teach these subjects to people.

Abelson is a Fellow of the IEEE and winner of the 1995 Taylor L. Booth Education Award given by IEEE Computer Society, cited for his continued contributions to the pedagogy and teaching of introductory computer science. He plays a leading role in educational technology at MIT as co-director of the MIT-Microsoft iCampus Research Alliance in Educational Technology and as co-chair of the MIT Council on Educational Technology. He is also one of the prime initiators of the MIT OpenCourseWare project.

Abelson's research at the MIT Artificial Intelligence Laboratory focuses on "amorphous computing," an effort to create programming technologies that can harness the power of the new computing substrates emerging from advances in microfabrication and molecular biology. He is also engaged in the interaction of law, policy, and technology as they relate to societal tensions sparked by the growth of the Internet. He initiated the MIT Computer Science Department's course on these topics, Ethics and Law on the Electronic Frontier, in 1994, and teaches it together with Daniel Weitzner.

Together with Gerald Sussman, Abelson developed MIT's introductory computer science subject, Structure and Interpretation of Computer Programs, a subject organized around the notion that a computer language is primarily a formal medium for expressing ideas about methodology, rather than just a way to get a computer to perform operations. This work, through a popular computer science textbook by Abelson and Gerald and Julie Sussman, videos of their lectures, and the availability on personal computers of the Scheme dialect of Lisp (used in teaching the course), has had a world-wide impact on university computer-science education.

Awards and Honors:

Phi Beta Kappa Visiting Scholar	2003–2004
IEEE Taylor Booth Award	1995
Elected Fellow of the IEEE	1994
MIT Class of 1922 Professorship	1994–
MIT Bose Award	1992
MIT MacVicar Faculty Fellow	1992–2002

Recent collaborators:

Tom Knight MIT
Peter Robinson Cambridge University, UK
Lawrence Lessig Stanford Law School
Jonathan Zittrain Harvard Law School
Daniel Weitzner World Wide Web Consortium

Recent PhD students supervised by Hal Abelson

Radhika Nagpal Harvard University
Ron Weiss Princeton University
Latanya Sweeney CMU
Daniel Coore University of the West Indies
Abelson has supervised the PhD theses of 13 students.

Curriculum Vitae for Timothy Berners-Lee

Tim Berners-Lee
MIT Computer Science and Artificial Intelligence Laboratory (CSAIL)
32 Vassar Street

MIT room 32-G524
Cambridge, MA 02139 USA
mailto:timbl@w3.org

Professional Preparation

- Oxford University, B.A. Physics, 1976

Appointments

- **Professor of Electronics and Computer Science**
2005 - present, Southampton University, UK. (ECS, dept., part time)
- **Senior Research Scientist**
2001 - present, Massachusetts Institute of Technology (MIT) Laboratory for Computer Science
- **Principal Research Scientist**
1995 - 2001, MIT Laboratory for Computer Science
- **3 Com Founders Chair**
1999 - present, MIT Laboratory for Computer Science
- **Director**
1994 - present, World Wide Web Consortium (W3C)
- **Research Scientist**
1994 - 1995, MIT Laboratory for Computer Science
- **Fellow/Staff Member**
1984 - 1986/1986 - 1994, CERN, European Center for Particle Physics Research
- **Director**
1984 - 1981, Image Computer Systems Ltd.
- **Private Consultant**
1980-1981
- **Software Engineer**
1978-1979, D.G.Nash Ltd.
- **Engineer (Assistant to Senior to Principal)**
1976-1978, Plessey Telecommunications Ltd.

Selected Publications

- Berners-Lee, T.J., et al, "World-Wide Web: Information Universe", Electronic Publishing: Research, Applications and Policy, 1992, p.4.
- Berners-Lee, T.J., et al, "The World Wide Web," Communications of the ACM, 1994, p. 8.
- T. Berners-Lee, L. Masinter, M. McCahill, "Universal Resource Locators (URL)", [RFC1738](#), 1994/12.
- T. Berners-Lee, D. Connolly "Hypertext markup Language - 2.0", [RFC1866](#), 1996/5.
- R. Fielding, J. Gettys, J. Mogul, H. Frystyk, L. Masinter, P. Leach, T. Berners-Lee "Hypertext Transfer Protocol HTTP 1.1", [RFC 2616](#), 1999/6.

- Tim Berners-Lee, Dan Connolly, Ralph R. Swick "[Web Architecture: Describing and Exchanging Data](#)", W3C Note, 1999/6-7.
- Berners-Lee, Tim, *Weaving the Web*, Harper San Francisco, 1999.
- Berners-Lee, Tim. and Hendler, James "[Publishing on the Semantic Web](#)", *Nature*, April 26 2001 p. 1023-1025.
- Berners-Lee, Tim; Hendler, James and Lassila, Ora "[The Semantic Web](#)", *Scientific American*, May 2001, p. 29-37.
- James Hendler, Tim Berners-Lee and Eric Miller, '[Integrating Applications on the Semantic Web](#)', *Journal of the Institute of Electrical Engineers of Japan*, Vol 122(10), October, 2002, p. 676-680.

Synergistic Activities

- Invented the World Wide Web, an internet-based hypermedia initiative for global information sharing, in 1989 while working at CERN and wrote the first web client (browser-editor) and server, and the original HTML, HTTP and URI specifications, in 1990.
- International World Wide Web Conference Steering Committee (1994 - present) and Program Committees (1993, 1994, 1995)
- World Economic Forum, invited sessions, 1997 and 1998, Davos, Switzerland.
- Congressional Internet Caucus Speakers Series, speech to members of Congress on the Semantic Web, the social implications of the use of the internet, June 2001.
- Radio and TV appearances to explain technologies and the Web, including NPR's *Science Friday* and *All Things Considered*.
- Board of Advisors, "Web Semantics: Science, Services and Agents on the World Wide Web" journal (2002 - present)

Collaborators and Other Affiliations

Fielding, Roy (Day Software), Gettys, Jim (Compaq), Hendler, James (University of Maryland), Frystyk-Nielsen, Henrik (Microsoft), Karger, David (CSAIL/MIT), Lassila, Ora (Nokia), Leach, Paul (Microsoft), Masinter, Larry (Adobe Systems), Mogul, Jeffrey (Hewlett-Packard), Stein, Lynn A. (Olin College)

JOAN FEIGENBAUM
Department of Computer Science, Yale University
P.O. Box 208285, New Haven, CT 06520-8285, U.S.A.
joan.feigenbaum@yale.edu, (203) 432-6432, <http://www.cs.yale.edu/homes/jf>

Professional preparation

Harvard University, Mathematics, BA, 1981
Stanford University, Computer Science, PhD, 1986

Appointments

Henry Ford II Professor of Computer Science, 1/06-present, Yale University
Professor, 7/00-12/05, Computer Science, Yale University, New Haven, CT 06520
Member of Research Staff, 1/96-6/00, AT&T Labs, Florham Park, NJ 07932
Department Head, 1/98-12/99, Algorithms and Distributed Data, AT&T Labs
Member of Technical Staff, 7/86-12/95, Bell Labs, Murray Hill, NJ 07974

Five publications most closely related to the project

D. Beaver, J. Feigenbaum, J. Kilian, and P. Rogaway, "Security with Low Communication Overhead," in *Advances in Cryptology – Crypto '90*, Lecture Notes in Computer Science, vol. 537, Springer, Berlin, 1991, pp. 62 - 76.

<http://www.springerlink.com/media/PF9DMXWHRN2QNNRLEFRAY/Contributions/V/M/Q/J/VMQJ92EPGBACRFE7.pdf>

D. Boneh, J. Feigenbaum, A. Silberschatz, and R. Wright, "PORTIA: Privacy, Obligations, and Rights in Technologies of Information Assessment," *Bulletin of the IEEE Technical Committee on Data Engineering* **27** (2004), pp. 10-18. <ftp://ftp.research.microsoft.com/pub/debull/A04mar/avi.ps>

D. Bergemann, T. Eisenbach, J. Feigenbaum, and S. Shenker, "Flexibility as an Instrument in Digital Rights Management," 2005 Workshop on Economics of Information Security (WEIS).

<http://www.cs.yale.edu/homes/jf/BEFS.pdf>

J. Feigenbaum, L. Fortnow, D. Pennock, and R. Sami, "Computation in a Distributed Information Market," *Theoretical Computer Science* **343** (2005), pp. 114-132.

<http://www.cs.yale.edu/homes/jf/FFPS.pdf>

J. Feigenbaum, Y. Ishai, T. Malkin, K. Nissim, M. Strauss, and R. Wright, "Secure Multiparty Computation of Approximations," to appear in *ACM Transactions on Algorithms*.

<http://www.cs.yale.edu/homes/jf/FIMNSW.pdf>

Five other publications

J. Feigenbaum and L. Fortnow, "Random-Self-Reducibility of Complete Sets," *SIAM Journal on Computing* **22** (1993), pp. 994 - 1005. <http://www.cs.yale.edu/homes/jf/FF.pdf>

J. Feigenbaum, "Games, Complexity Classes, and Approximation Algorithms," in *Proceedings of the International Congress of Mathematicians, volume III; Invited Lectures*, Documenta Mathematica, Journal der Deutschen Mathematiker-Vereinigung, 1998, pp. 429-439.

<http://www.cs.yale.edu/homes/jf/F-ICM.pdf>

J. Feigenbaum and S. Shenker, "Distributed Algorithmic Mechanism Design: Recent Results and Future Directions," in *Proceedings of the 6th International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications (DIALM)*, ACM Press, New York, 2002, pp. 1543. <http://www.cs.yale.edu/homes/jf/FS.pdf>

D. Bergemann, J. Feigenbaum, S. Shenker, and J. Smith, "Towards an Economic Analysis of Trusted Systems," 2004 Workshop on Economics of Information Security (WEIS).

<http://www.dtc.umn.edu/weis2004/feigenbaum.pdf>

J. Feigenbaum, S. Kannan, A. McGregor, S. Suri, and J. Zhang, "Graph Distances in the Streaming Model: The Value of Space," in *Proceedings of the 16th Symposium on Discrete Algorithms (SODA)*, ACM/SIAM, New York/Philadelphia, 2005, pp. 745 - 754.

<http://www.cs.yale.edu/homes/jf/FKMSZ2.pdf>

Synergistic activities

Program-Committee Chair or Co-Chair, *Crypto* 1991; *ACM Workshop on Digital Rights Management* 2002; *ACM Conference on Electronic Commerce* 2004; *DIMACS Workshops on Trust-Management and Public-Key Infrastructure* 1996 and *Management of Digital IP* 2000

Editor-in-chief, *Journal of Cryptology*, 1997-2002

Co-Chair, DIMACS Special Focus *Massive Data Sets*, 1997-1999

Co-Chair, DIMACS Special Focus on *Next-Generation Networks*, 2000-2003

Yale PI, PORTIA Project on Sensitive Data in a Wired World, 2003-present

Awards

Invited lecturer, International Congress of Mathematicians, 1998

ACM Fellow, 2001-present

Research Interests

Privacy and security, massive-data-set algorithms, and incentive-compatible distributed computation

Collaborators and other affiliations

Co-authors and co-PIs: A. Archer (AT&T Labs), J. Aspnes (Yale), J. Balkin (Yale), D. Bergemann (Yale), D. Boneh (Stanford), G. Crabb (US Secret Service), C. Dwork (Microsoft), T. Eisenbach (Univ. Munich), S. Forrest (Univ. New Mexico), L. Fortnow (Univ. Chicago), H. Garcia-Molina (Stanford), B. Grosz (MIT), J. Halpern (Cornell), S. Hawala (US Census Bureau), Y. Ishai (Technion), S. Jha (Univ. Wisconsin), R. Kannan (Yale), S. Kannan (Univ. Pennsylvania), D. Karger (MIT), A. Keromytis (Columbia), B. Knutsson (Swedish Royal Inst. of Technology), A. Krishnamurthy (Univ. Washington), B. LaMacchia (Microsoft), I. Lee (Univ. Pennsylvania), N. Li (Purdue), P. Lincoln (SRI), T. Malkin (Columbia), K. McCurley (Google), A. McGregor (Univ. Pennsylvania), P. Miller (Yale), V. Mirrokni (Microsoft), J. Mitchell (Stanford), M. Mitzenmacher (Harvard), J. Morris (CDT), R. Motwani (Stanford), H. Nissenbaum (NYU), K. Nissim (Ben-Gurion Univ.), C. Papadimitriou (UC Berkeley), D. Parkes (Harvard), D. Pennock (Yahoo), B. Pinkas (Haifa Univ.), T. Roughgarden (Stanford), R. Ryger (Yale), F. Saint-Jean (Yale), R. Sami (Univ. Michigan), A. Scedrov (Univ. Pennsylvania), A. Schaeffer (DHHS/NIH), M. Schapira (Hebrew Univ.), D. Schutzer (Citigroup), S. Shenker (ICSI and UC Berkeley), V. Shmatikov (Univ. Texas), J. Zhang (SRI), S. Zhong (SUNY Buffalo)

PhD Advisor: A. C. Yao (currently at Tsinghua Univ.)

Postdoctoral Advisor: N/A

PhD Students (9 total): R. Dakdouk (current, Yale), A. Johnson (current, Yale), N. Li (PhD 2000, NYU; currently at Purdue), V. Ramachandran (PhD 2005, Yale; currently at Stevens Inst. of Tech.), R. Ryger (current, Yale), F. Saint-Jean (current, Yale), R. Sami (PhD 2003, Yale; currently at Univ. Michigan), J. Zhang (PhD 2005, Yale; currently at SRI), S. Zhong (PhD 2004, Yale; currently at SUNY Buffalo)

Postdocs (3 total): M. Elkin (2003-2004, Yale; currently at Ben-Gurion Univ.), F. Esponda (current, Yale), N. Kozlovski (2004-2005, Yale; currently at the Israeli Ministry of Justice)

Chris Hanson

Education

MIT S.B. (Computer Science) 1980

Professional Appointments

Principal Research Scientist, MIT Computer Science and Artificial Intelligence Laboratory
(1983 – present)

Research Scientist, Artificial Intelligence Laboratory, Massachusetts Institute of Technology
(1982 – 1993)

Software Engineer, Data Translation Inc.
(1980 – 1982)

Selected publications relevant to this proposal:

Transparent Accountable Data Mining: New Strategies for Privacy Protection, Weitzner, Abelson, Berners-Lee, et al., [MIT CSAIL Technical Report MIT-CSAIL-TR-2006-007 \(27 January 2006\)](#)

Selected other publications:

The Revised^f Report on the Algorithmic Language Scheme, J. Rees and W. Clinger, eds., in *Lisp Pointers* 4(3), ACM, 1991.

IEEE Standard for the Scheme Programming Language—IEEE Std 1178-1990, David Bartley, Chris Hanson, and James Miller, eds., IEEE Computer Society, December, 1990.

Efficient Stack Allocation for Tail-Recursive Languages, Chris Hanson, in *Proceedings, ACM Conference on Lisp and Functional Programming*, Nice, France, June 1990.

The Scheme-81 Architecture—System and Chip, John Batali, Edmund Goodhue, Chris Hanson, Howie Shrobe, Richard M. Stallman, and Gerald Jay Sussman, *Proceedings, Conference on Advanced Research in VLSI*, MIT, Cambridge MA, January 1982.

Synergistic Activities:

Software developer on the [TAMI project](#).

Chief architect of the [MIT/GNU Scheme](#) programming environment.

Member of the [Debian project](#), which develops and distributes a free operating system.

Collaborators:

Prof. Harold Abelson (MIT)

Tim Berners-Lee (MIT)

Prof. Gerald Jay Sussman (MIT)

Daniel J. Weitzner (MIT).

Recent Teaching Experience:

MIT 6.891: [Adventures in Advanced Symbolic Programming](#), with Gerald Jay Sussman, Spring 2006.

Professor James A. Hendler

University of Maryland, College Park

Jim Hendler is a Professor at the University of Maryland and the Director of Semantic Web and Agent Technology at the Maryland Information and Network Dynamics Laboratory. He has joint appointments in the Department of Computer Science and the Institute for Advanced Computer Studies and is an affiliate of the Institute for Systems Research. He has authored over 150 technical papers in the areas of artificial intelligence, Semantic Web, agent-based computing and high performance processing. Hendler was the recipient of a 1995 Fulbright Foundation Fellowship, is a former member of the US Air Force Science Advisory Board, and is a Fellow of the American Association for Artificial Intelligence. He is also the former Chief Scientist of the Information Systems Office at the US Defense Advanced Research Projects Agency (DARPA), was awarded a US Air Force Exceptional Civilian Service Medal in 2002, and is actively involved in the Semantic Web Activity at the W3C. He is currently a member of the Board of Reviewing Editors for *Science* and the Editor-in-Chief for *IEEE Intelligent Systems*.

Current Academic Activities:

Director, Semantic Web and Agent Technology, U. Maryland Information and Network Dynamics Laboratory.

Editor-in-Chief, *IEEE Intelligent Systems*

Board of Reviewing Editors, *Science*

Chair of Advisory Board, Web Semantics Journal (Elsevier Publishing)

Associate Editor or Ed Board Member:, *ACM Transactions on Internet Technology*, *Journal of Experimental and Theoretical Artificial Intelligence*, *Artificial Intelligence Journal*, *IEEE Intelligent Systems*, *Autonomous Robotics*, *Electronic Transactions on AI: Planning and Scheduling*, *Journal on Autonomous Agents and Multiagent System*, *Journal of Cognitive Theories and Systems (Electronic Journal)*

Chair of the Advisory Board of the Protégé project, Stanford, 2003-pres.

Chair, Conference Committee, American Association for Artificial Intelligence

Award and Honors

DIA DDL Senior Advisory Group, 1/06-pres.

Board of Reviewing Editors, *Science*, 1/2005 – pres.

Robert Engelmores Memorial Prize, American Association of Artificial Intelligence, 2005

National Academy Navy Studies Board (ad hoc member), 2005

NASA Earth Science Activity Technical Advisory Board, 2003-2005

US Air Force Exceptional Civilian Service Medal, 10/02.

Expository Writing Award, American Association of Artificial Intelligence, Aug 2000

Air Force Science Advisory Board, 1998-2002.

Fellow, American Association for Artificial Intelligence, 1999.

Invention of the Year Award, PARKA-DB™, Univ of Maryland, 1997.

Fulbright Fellowship, Senior Researcher to Israel, CIES/USIEF, 1995-1996.

Five Related Publications

A. Kalyanpur, B. Parsia, E. Sirin, B. Cuenca-Grau, and J.Hendler. Swoop - a web ontology editing browser. *Journal of Web Semantics*, 4(1), 2005.

U. Kuter, E. Sirin, B. Parsia, D. Nau, and J. Hendler. Information gathering during planning for web service composition. *Journal of Web Semantics*, 3(2), 2005.

J. Golbeck and J.Hendler. Inferring trust relationships in web-based social networks. *ACM Transactions on Internet Technology*, (to appear).

G Jiang, G Cybenko, and JHendler. Semantic interoperability and information fluidity. *Intl Journal of Cooperative Information Systems*, 2005.

A. Kalyanpur, B.Parsia, E. Sirin, and J. Hendler. Debugging unsatisfiable classes in owl ontologies. *Journal of Web Semantics*, 3(4), 2005.

Five Other Assorted Publications

D. De Roure and J. Hendler, E-science: The Grid and the Semantic Web, *IEEE Intelligent Systems*, February, 2004.

J. Hendler, Science and the Semantic Web, *Science*, Vol 299, Oct 24, 2003.

J. Hendler, Agents and the Semantic Web, *IEEE Intelligent Systems*, 16(2), March/April, 2001.

T. Berners-Lee, J. Hendler and O. Lassila, The Semantic Web: When the Internet gets Smart, *Scientific American*, 284(5), May 2001

Translations: Det semantiske Web (Dutch) in: *Nar Nettet Aendrer Verden*, M. Jensen, M. Pedersen, and A. Talbro (eds), Borsens Verlag: Denmark, Sept 2001; "Mein Computer versteht mich" *Spektrum der Wissenschaft* (German, August 2001); "La Red semántica" from *Investigación y Ciencia* (Spanish, July 2001); "Il Web semantico" from *Le Scienze* (Italian, May 2001), "Sieç Sematyczna" from *Swiat Nauki* (Polish, 7/2001)

J. Hendler, High Performance Artificial Intelligence, *Science*, Vol 265, Aug 12, 1994.

Collaborators and Co-Editors

From 1998-2001, Dr. Hendler worked as a government funding agent for AFOSR and DARPA. During that time he funded over 100 researchers in the areas of artificial intelligence, agents, Semantic Web, and Software Development. A complete list is available on demand, but the NSF has previously ruled there is no remaining conflict of interest with these researchers.

Coauthors/Coeditors since 2001 include: Tim Berners-Lee, Ronald Brachman, Rodney Brooks, George Cybenko, David de Roure, Ed Feigenbaum, Dieter Fensel, Tim Finin, Gilberto Fragoso, Carole Goble, Frank Hartel, Wendy Hall, Ian Horrocks, Yarden Katz, Ora Lassila, Yiannis Labrou, Henry Lieberman, Ryu Masuoka, David McCombs, Eric Miller, Dana Nau, Bijan Parsia, Nigel Shadbolt, Peter Patel-Schneider, Raj Reddy, Joel Sachs, Hava Siegelmann, VS Subrahmanian, Wolfgang Wahlster, and Daniel Weitzner.

Graduate Advisor: Professor Eugene Charniak, Brown University.

Gerald Jay Sussman

Education:

MIT S.B. (Mathematics)	1968
MIT Ph.D. (Mathematics)	1973

Professional Appointments:

1991–present	Matsushita Professor of Electrical Engineering MIT
1984–1991	Full Professor of Electrical Engineering MIT
1977–1984	Associate Professor of Electrical Engineering MIT
1973–1977	Assistant Professor of Electrical Engineering MIT

Selected publications relevant to this proposal:

1. Ron Weiss, Thomas F. Knight, and Gerald Jay Sussman, “Genetic Process Engineering,” in Cellular Computing, Martyn Amos editor, pp.43–73, Oxford University Press, 2004.
2. “Amorphous Computing,” Harold Abelson, Don Allen, Daniel Coore, Chris Hanson, George Homsy, Thomas F. Knight Jr., Radhika Nagpal, Erik Rauch, Gerald Jay Sussman, and Ron Weiss, in Communications of the ACM, 43, 5, May 2000.
3. “Cellular Gate Technology,” Thomas F. Knight and Gerald Jay Sussman, Proc. UMC98, First International Conference on Unconventional Models of Computation, Auckland, NZ, January 1998.
4. Structure and Interpretation of Computer Programs, Hal Abelson, Gerald Jay Sussman and Julie Sussman, MIT Press and McGraw-Hill, 1985, (published translations in French, Polish, Chinese, Japanese, and German). Second Edition, 1996.
5. “Intelligence in Scientific Computing,” Hal Abelson, M. Eisenberg, M. Halfant, J. Katzenelson, E. Sacks, G.J. Sussman, J. Wisdom, K. Yip, CACM, 32, no. 5, May 1989.

Selected other publications:

1. Structure and Interpretation of Classical Mechanics, Gerald Jay Sussman and Jack Wisdom, with Meinhard Mayer, MIT Press, 2001.
2. “The first report on Scheme revisited,” Gerald Jay Sussman and Guy L. Steele Jr., Higher-Order and Symbolic Computation, 11, No.4, pp. 399-404, 1998.
3. “Spin-induced Orbital Precession and its Modulation of the Gravitational Waveforms from Merging Binaries,” T.A. Apostolatos, C. Cutler, G.J. Sussman, and K.S. Thorne, Phys. Rev. D., 15 June 1994.
4. “Increasing the Compressive Strength of a Column via Active Control”, A. Berlin and G.J. Sussman, Proceedings of the Third International Conference on Adaptive Structures, Oct 1992.
5. “Chaotic Evolution of the Solar System,” Gerald Jay Sussman and Jack Wisdom, Science, 257, 3 July 1992.

Synergistic Activities:

Sussman’s research has centered on understanding the problem-solving strategies used by scientists and engineers, with the goals of automating parts of the process and formalizing it to provide more

effective methods of science and engineering education. Sussman's contributions include problem solving by debugging almost-right plans, propagation of constraints applied to electrical circuit analysis and synthesis, dependency-based explanation and dependency-based backtracking, and various language structures for expressing problem-solving strategies. Sussman and his former student, Guy L. Steele Jr., invented the Scheme programming language in 1975. Sussman is a coauthor (with Hal Abelson and Julie Sussman) of the introductory computer science textbook used at M.I.T. The textbook, "Structure and Interpretation of Computer Programs," has been translated into many languages. As a result of this and other contributions to computer-science education, Sussman received the ACM's Karl Karlstrom Outstanding Educator Award, and the Amar G. Bose award. Sussman was the principal designer of the Digital Orrery, a machine designed to do high-precision integrations for orbital-mechanics experiments. The Orrery was designed and built by a few people in a few months, using AI-based simulation and compilation tools. Using the Digital Orrery, Sussman has worked with Jack Wisdom to discover numerical evidence for chaotic motions in the outer planets. The Digital Orrery is now retired at the Smithsonian Institution in Washington DC.

Awards and Honors:

2004	Elected Fellow of the American Association for the Advancement of Science
2004	Elected Fellow of the New York Academy of Sciences
2002	Elected Fellow of the Institute of Electrical and Electronics Engineers
2000	Elected Member of the National Academy of Engineering
1996	Elected Fellow of the American Academy of Arts and Sciences
1994	Elected Founding Fellow of the ACM
1992	Amar G. Bose award for Engineering Education
1991	ACM Karl Karlstrom Outstanding Educator Award
1990	Elected Founding Fellow of the AAAI

Recent collaborators:

Harold Abelson MIT
Drew Endy MIT
Tom Knight MIT
Peter Robinson Cambridge University, UK
Ron Weiss Princeton
Jack Wisdom MIT

Graduate Advisors of Gerald Jay Sussman

Marvin L. Minsky MIT
Seymour A. Papert MIT

Recent PhD students supervised by Gerald Jay Sussman

Daniel Coore University of the West Indies
Radhika Nagpal Harvard University
Ron Weiss Princeton University
Erik Rauch Princeton University
Attila Kondacs not yet placed

Sussman has supervised the PhD theses of 35 students.

Daniel Jacob Weitzner

Principal Research Scientist
MIT Computer Science and Artificial Intelligence Laboratory
<http://www.w3.org/People/Weitzner.html>

Professional Preparation

Swarthmore College	BA (Philosophy)	1985
SUNY Buffalo Law School	JD (Cum Laude)	1992

Appointments

Co-founder and Principal Investigator, MIT CSAIL Decentralized Information Group
(January 2005 – present)

Visiting Professor, University of Southampton, Electronics & Computer Science (March 2006 - present)

Principal Research Scientist, MIT Computer Science and Artificial Intelligence Laboratory (2003 – present)

Technology and Society Domain Lead, World Wide Web Consortium (1998 – present)

Co-founder & Deputy Director, Center for Democracy and Technology (1994 - 1998)

Deputy Policy Director, Electronic Frontier Foundation (1991 – 1994)

Publications

Closely Related:

Weitzner, Abelson, Berners-Lee, et al., "Transparent Accountable Data Mining: New Strategies for Privacy Protection", MIT CSAIL Technical Report MIT-CSAIL-TR-2006-007 (27 January 2006). (<http://www.w3.org/2006/01/tami-privacy-strategies-aaai.pdf>)

Lalana Kagal, Tim Berners-Lee, Dan Connolly, and Daniel Weitzner, [Using Semantic Web Technologies for Policy Management on the Web](#), [21st National Conference on Artificial Intelligence](#) (AAAI), July 16 - 20, 2006.

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, 2005.

Ackerman, M., Darrell, T., & Weitzner, D. J. (2001). [Privacy in context](#). *Human-Computer Interaction*, 16, pp. 167-176.

Testimony before the US Senate Commerce Committee, [Hearing on Online Privacy](#), 25 May 2000

Berman, Jerry and Daniel J. Weitzner. "Directing Policy-Making Around the Net's Metaphor." *Communications of the ACM* v.40 (February 1997): pp.83-84

Jerry Berman & Daniel Weitzner, [Abundance and User Control: Renewing the Democratic Heart of the First Amendment in the Age of Interactive Media](#), 104 *Yale L.J.* 1619 (1995)

Others:

Testimony before the US Federal Trade Commission and US Department of Justice Joint Hearings on Competition and Intellectual Property Law and Policy in the Knowledge-Based Economy: Standards and Intellectual Property: Licensing Terms (18 April 2002)

Berman, J., & Weitzner, Daniel "Technology and Democracy." [Social Research](#) 64 Fall 1997: 1313-19

Kapor, Mitch and Weitzner, Daniel (1994). "Social and Industrial Policy for Public Networks: Visions for the Future". Harasim and Walls, eds. *Global Networks: Computers and International Communication*. Oxford University Press. Oxford.

Synergistic Activities

Member, National Academy of Sciences Study Committee on [Technical and Privacy Dimensions of Information for Terrorism Prevention and Other National Goals](#) (April 2006 – October 2007)

Founding Board Member, Software Freedom Law Center (January 2005 – present)

Editorial Board, *Foundations and Trends in Web Science*

Editorial Board, [Journal of Privacy Technology](#)

Chair, W3C Patent Policy Working Group (also Editor of W3C's Royalty-Free Patent Policy)

Member, National Academy of Sciences Study Committee on [Authentication Technologies and their Privacy Implications](#) (March 2000- November 2003)

Member, Pew Charitable Trusts Democracy Online Task Force (April 1999 – 2002)

Founding Board Member, Center for Democracy and Technology (1994 - present)

Collaborators and other Affiliations

Collaborators:

T. Berners-Lee (MIT)
J. Hendler (U. Maryland)
M. Ackerman (Michigan)

Thesis Advisor and Postgraduate-Scholar Sponsor:

Lalana Kagal (Post Doctoral Fellow, MIT CSAIL: 2004 – present)
Simson Garfinkel (PhD dissertation committee)
Latanya Sweeney (PhD dissertation committee)

Selected Invited Lectures

Keynote, "Broken Links on the Web: Local Laws and the Global Free Flow of Information," at the 15th Annual World Wide Web Conference, Edinburgh, Scotland, 26 May 2006.

Keynote "Semantic Web Public Policy Challenges: Privacy, Provenance, Property and Personhood" at the "4th International Semantic Web Conference" (9 November 2005)

DHS Privacy Advisory Committee, Testimony on the impact of Semantic Web technologies on privacy, civil liberties, and homeland security. 15 July 2005

"The Transparency Paradox," UC Berkeley School of Information Management & Systems: Distinguished Lecture Series (10 November 2004)

Keynote, Human Computer Interaction Consortium, "Can Transparency Save Us?: Design Goals for More Sociable Information Spaces" (7 February 2003)

Crypto 2001 conference, "Privacy, Authentication & Identity: A recent history of cryptographic struggles for freedom"

Recent Teaching Experience

MIT 6.805/6.806/STS085: Ethics and Law on the Electronic Frontier: Copyright & Digital Rights Management Technologies, with H. Abelson (Fall 2004, 2003, 2002, 2000, 1999, 1998)

Section 7: Offeror Statement of Work

1.0 OBJECTIVE

1.1 The objective of this project is to design Web-scale information architecture that enables end-to-end semantic accountability with respect to the laws, rules and procedures that apply to large scale information sharing environments. End-to-end semantic accountability is a property of a networked information environment that provides for efficient and scalable assessment of rules compliance across all uses of information without unduly impeding sharing or use of the information.

2.0 SCOPE

2.1 The scope of this effort is to develop technology that will enables the Intelligence Community, others in the information sharing environment, and authorized regulators to assess policy compliance of information sharing and usage activities at the appropriate time, according to the rules relevant to that information.

3.0 BACKGROUND

3.1 There is an urgent need for transparency and accountability in a variety of information applications that depend upon sensitive, personal information, both in the public and the commercial sector. Attempts to address issues of personal privacy in a world of computerized databases and information networks typically proceed from the perspective of preventing or controlling *access* to information. We argue that this perspective has become inadequate and obsolete, overtaken by the effortless sharing and copying data, and the ease of aggregating and searching across multiple data bases to reveal private information from public sources.

3.2 The need for end-to-end semantic accountability is illustrated by considering the unique policy and trustworthiness requirements of an interagency information-sharing environment that links data from classified intelligence sources, sensitive civilian law enforcement records, Federal and State agency records, and commercial data sources, along with publicly available data such as that residing on the World Wide Web.

3.3 The key threats against which we will design our end-to-end accountability infrastructure center around the dual threat that information may be shared too widely and misused, together with the mirror image threat that uncertainty about whether particular transfers or uses of information are permitted may *inhibit* what would be both legal and beneficial sharing and analysis.

4.0 TASKS/TECHNICAL REQUIREMENTS:

4.1 The contractor shall accomplish the following:

4.1.1 Develop scenarios describing information sharing activities along with the applicable rules against with accountability should be judged by the accountability appliances and browser.

4.1.2 Develop a description and formal specification of a Data Purpose Algebra which express rules describing how data can be used.

4.1.3 Develop a prototype Accountability Appliance which should include:

4.1.3.1 Client-side http proxy, constructing access proofs and logging transactions in a secure manner

4.1.3.2 A Web server module checking and acting on access requests (using a proof checker), providing secure logging of all information access and usage transactions and depositing these transaction records into a TMS.

4.1.3.3 General accountability browser, providing users the ability to evaluate whether or not a given use of a collection of information complies with the relevant rules and policies.

4.1.4 Evaluation reports on the output of this project (each item described in 4.1.3) which shall include an assessment of the scalability of the accountability architecture, expressivity of the policy languages, and resilience against identified threats.

4.1.5. Report on progress as required by the program managers, including:

4.1.5.1 Monthly status reports (including brief financial status), briefing charts, a final technical report.

4.1.5.2. Oral presentations for program reviews which shall include the status of the technical progress made to date in the performance of the contract.