COVER SHEET FOR PROPOSAL TO THE NATIONAL SCIENCE FOUNDATION

PROGRAM ANNOUNCEMENT/SOLICITATION NO./CLOSING DATE/if not in response to a program announcement/solicitation enter NSF 07-140	FOR NSF USE ONLY
NSF 07-601 **01/07/08**	**NSF PROPOSAL NUMBER**

FOR CONSIDERATION BY NSF ORGANIZATION UNIT(S)   (Indicate the most specific unit known, i.e. program, division, etc.)

OCI - DATA INTEROPERABILITY NETWORKS

DATE RECEIVED	NUMBER OF COPIES	DIVISION ASSIGNED	FUND CODE	DUNS# (Data Universal Numbering System)	FILE LOCATION
				001425594	

EMPLOYER IDENTIFICATION NUMBER (EIN) OR TAXPAYER IDENTIFICATION NUMBER (TIN)	SHOW PREVIOUS AWARD NO. IF THIS IS ☐ A RENEWAL ☐ AN ACCOMPLISHMENT-BASED RENEWAL	IS THIS PROPOSAL BEING SUBMITTED TO ANOTHER FEDERAL AGENCY? YES ☐ NO ☒ IF YES, LIST ACRONYM(S)
042103594		

NAME OF ORGANIZATION TO WHICH AWARD SHOULD BE MADE	ADDRESS OF AWARDEE ORGANIZATION, INCLUDING 9 DIGIT ZIP CODE
Massachusetts Institute of Technology	**77 MASSACHUSETTS AVE** **Cambridge, MA 02139-4307**

AWARDEE ORGANIZATION CODE (IF KNOWN)

0021782000

NAME OF PERFORMING ORGANIZATION, IF DIFFERENT FROM ABOVE	ADDRESS OF PERFORMING ORGANIZATION, IF DIFFERENT, INCLUDING 9 DIGIT ZIP CODE

PERFORMING ORGANIZATION CODE (IF KNOWN)

IS AWARDEE ORGANIZATION (Check All That Apply) (See GPG II.C For Definitions)	☐ SMALL BUSINESS ☐ FOR-PROFIT ORGANIZATION	☐ MINORITY BUSINESS ☐ WOMAN-OWNED BUSINESS	☒ IF THIS IS A PRELIMINARY PROPOSAL THEN CHECK HERE

TITLE OF PROPOSED PROJECT **DataNet Preliminary Proposal: DataSpace Project**

REQUESTED AMOUNT	PROPOSED DURATION (1-60 MONTHS)	REQUESTED STARTING DATE	SHOW RELATED PRELIMINARY PROPOSAL NO. IF APPLICABLE
$ **2**	**60** months	**06/01/08**	

CHECK APPROPRIATE BOX(ES) IF THIS PROPOSAL INCLUDES ANY OF THE ITEMS LISTED BELOW

☐ BEGINNING INVESTIGATOR (GPG I.G.2)
☐ DISCLOSURE OF LOBBYING ACTIVITIES (GPG II.C)
☐ PROPRIETARY & PRIVILEGED INFORMATION (GPG I.D, II.C.1.d)
☐ HISTORIC PLACES (GPG II.C.2.j)
☐ SMALL GRANT FOR EXPLOR. RESEARCH (SGER) (GPG II.D.1)
☐ VERTEBRATE ANIMALS (GPG II.D.5) IACUC App. Date _____
   PHS Animal Welfare Assurance Number _____

☐ HUMAN SUBJECTS (GPG II.D.6) Human Subjects Assurance Number _____
   Exemption Subsection _____ or IRB App. Date _____
☐ INTERNATIONAL COOPERATIVE ACTIVITIES: COUNTRY/COUNTRIES INVOLVED
   (GPG II.C.2.j)
   _____
☐ HIGH RESOLUTION GRAPHICS/OTHER GRAPHICS WHERE EXACT COLOR REPRESENTATION IS REQUIRED FOR PROPER INTERPRETATION (GPG I.G.1)

PI/PD DEPARTMENT **Information Technology**	PI/PD POSTAL ADDRESS **50 Memorial Drive** **Room E53-321** **Cambridge, MA 02139** **United States**
PI/PD FAX NUMBER **617-253-3321**	

NAMES (TYPED)	High Degree	Yr of Degree	Telephone Number	Electronic Mail Address
PI/PD NAME **Stuart E Madnick**	**PhD**	**1972**	**617-253-6671**	**smadnick@mit.edu**
CO-PI/PD **Harold Abelson**	**PhD**	**1973**	**617-253-5856**	**HAL@MIT.EDU**
CO-PI/PD **Jerrold M Grochow**	**DPhil**	**1974**	**617-253-1000**	**jgrochow@mit.edu**
CO-PI/PD **MacKenzie Smith**	**MLS**	**1986**	**617-253-1000**	**kenzie@mit.edu**
CO-PI/PD				

CERTIFICATION PAGE

Certification for Authorized Organizational Representative or Individual Applicant:

By signing and submitting this proposal, the Authorized Organizational Representative or Individual Applicant is: (1) certifying that statements made herein are true and complete to the best of his/her knowledge; and (2) agreeing to accept the obligation to comply with NSF award terms and conditions if an award is made as a result of this application. Further, the applicant is hereby providing certifications regarding debarment and suspension, drug-free workplace, and lobbying activities (see below), nondiscrimination, and flood hazard insurance (when applicable) as set forth in the NSF Proposal & Award Policies & Procedures Guide, Part I: the Grant Proposal Guide (GPG) (NSF 07-140). Willful provision of false information in this application and its supporting documents or in reports required under an ensuing award is a criminal offense (U. S. Code, Title 18, Section 1001).

Conflict of Interest Certification

In addition, if the applicant institution employs more than fifty persons, by electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative of the applicant institution is certifying that the institution has implemented a written and enforced conflict of interest policy that is consistent with the provisions of the NSF Proposal & Award Policies & Procedures Guide, Part II, Award & Administration Guide (AAG) Chapter IV.A; that to the best of his/her knowledge, all financial disclosures required by that conflict of interest policy have been made; and that all identified conflicts of interest will have been satisfactorily managed, reduced or eliminated prior to the institution's expenditure of any funds under the award, in accordance with the institution's conflict of interest policy. Conflicts which cannot be satisfactorily managed, reduced or eliminated must be disclosed to NSF.

Drug Free Work Place Certification

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative or Individual Applicant is providing the Drug Free Work Place Certification contained in Exhibit II-3 of the Grant Proposal Guide.

Debarment and Suspension Certification   (If answer "yes", please provide explanation.)

Is the organization or its principals presently debarred, suspended, proposed for debarment, declared ineligible, or voluntarily excluded
from covered transactions by any Federal department or agency?   Yes ☐   No ☒

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative or Individual Applicant is providing the Debarment and Suspension Certification contained in Exhibit II-4 of the Grant Proposal Guide.

Certification Regarding Lobbying

The following certification is required for an award of a Federal contract, grant, or cooperative agreement exceeding $100,000 and for an award of a Federal loan or a commitment providing for the United States to insure or guarantee a loan exceeding $150,000.

Certification for Contracts, Grants, Loans and Cooperative Agreements

The undersigned certifies, to the best of his or her knowledge and belief, that:

(1) No federal appropriated funds have been paid or will be paid, by or on behalf of the undersigned, to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with the awarding of any federal contract, the making of any Federal grant, the making of any Federal loan, the entering into of any cooperative agreement, and the extension, continuation, renewal, amendment, or modification of any Federal contract, grant, loan, or cooperative agreement.

(2) If any funds other than Federal appropriated funds have been paid or will be paid to any person for influencing or attempting to influence an officer or employee of any agency, a Member of Congress, an officer or employee of Congress, or an employee of a Member of Congress in connection with this Federal contract, grant, loan, or cooperative agreement, the undersigned shall complete and submit Standard Form-LLL, "Disclosure of Lobbying Activities," in accordance with its instructions.

(3) The undersigned shall require that the language of this certification be included in the award documents for all subawards at all tiers including subcontracts, subgrants, and contracts under grants, loans, and cooperative agreements and that all subrecipients shall certify and disclose accordingly.

This certification is a material representation of fact upon which reliance was placed when this transaction was made or entered into. Submission of this certification is a prerequisite for making or entering into this transaction imposed by section 1352, Title 31, U.S. Code. Any person who fails to file the required certification shall be subject to a civil penalty of not less than $10,000 and not more than $100,000 for each such failure.

Certification Regarding Nondiscrimination

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative is providing the Certification Regarding Nondiscrimination contained in Exhibit II-6 of the Grant Proposal Guide.

Certification Regarding Flood Hazard Insurance

Two sections of the National Flood Insurance Act of 1968 (42 USC §4012a and §4106) bar Federal agencies from giving financial assistance for acquisition or construction purposes in any area identified by the Federal Emergency Management Agency (FEMA) as having special flood hazards unless the:

(1)   community in which that area is located participates in the national flood insurance program; and
(2)   building (and any related equipment) is covered by adequate flood insurance.

By electronically signing the NSF Proposal Cover Sheet, the Authorized Organizational Representative or Individual Applicant located in FEMA-designated special flood hazard areas is certifying that adequate flood insurance has been or will be obtained in the following situations:

(1)   for NSF grants for the construction of a building or facility, regardless of the dollar amount of the grant; and
(2)   for other NSF Grants when more than $25,000 has been budgeted in the proposal for repair, alteration or improvement (construction) of a building or facility.

AUTHORIZED ORGANIZATIONAL REPRESENTATIVE	SIGNATURE	DATE
NAME		
TELEPHONE NUMBER	ELECTRONIC MAIL ADDRESS	FAX NUMBER

*SUBMISSION OF SOCIAL SECURITY NUMBERS IS VOLUNTARY AND WILL NOT AFFECT THE ORGANIZATION'S ELIGIBILITY FOR AN AWARD. HOWEVER, THEY ARE AN INTEGRAL PART OF THE INFORMATION SYSTEM AND ASSIST IN PROCESSING THE PROPOSAL. SSN SOLICITED UNDER NSF ACT OF 1950, AS AMENDED.

DataNet Preliminary Proposal: MIT DataSpace – Project Summary

PI: Stuart Madnick (MIT); Co-PI's: Hal Abelson (MIT), Jerrold Grochow (MIT), MacKenzie Smith (MIT); Potential Senior Personnel: Timothy Berners-Lee (W3C, WSRI), Pat Dreher (RENCI), John Erickson (HP Labs), Geneva Henry (Rice University), Mei Hsu (HP Labs), Henry Jacoby (MIT), David Karger (MIT), Isaac Kohane (Harvard), Michele Kimpton (DSpace Foundation), Tom Malone (MIT), Alexa McCray (Harvard), Jill Mesirov (Broad Institute), Ronald Prinn (MIT), Michael Siegel (MIT), Michael Stonebraker (MIT), Tyler Walters (Georgia Tech), Danny Weitzner (W3C, WSRI), John Wilbanks (Science Commons).

Summary and Vision: Web technology brought tremendous efficiency gains for commerce, yet the world of scientific research has failed to fully leverage all its capabilities. As a result scientists duplicate research and miss opportunities for discovery, collaboration and translation of research into public goods. The DataSpace Project will bring these gains to science by providing a dramatically new approach to data management and long-term curation that accommodates multiple, heterogeneous data from a variety of distributed locations, and supports research across diverse disciplines and modalities, enabling investigators to easily access and aggregate data of known quality and provenance. It will build on proven technology and business models while bringing to bear MIT's and our partner organizations' best research. To encourage sustainability and collaboration, organizations producing research data will be able to take responsibility for long-term stewardship of their data as part of a global network with only modest investment and expertise.

Intellectual Merit: the MIT DataSpace Project implements this model by
- Working closely with researchers and data in multiple scientific research domains to insure *relevance*, and *interoperable, interdisciplinary* solutions to data access and management. Initial domains are life science (e.g. bioinformatics and biomedical image data) and energy/environment (e.g. climate change data). Additional domains (e.g. high energy physics and plant biology) will be included later to test interdisciplinarity and interoperability of the system design.
- Building on our experience with existing *distributed infrastructure* for digital archiving and long-term preservation, but extending it to support the scale and complexity of research data. The project will build on the DSpace open source digital archive platform and its current federation of almost 300 research-generating organizations world wide.
- Leveraging a large number of MIT and external collaborators who have *extensive and proven expertise* in key areas: data integration, quality and interoperability, representation and visualization, database and storage technology, federated system design and policy management, long-term preservation, public outreach and education, and business planning and management for organizations.

Broader Impact: the MIT DataSpace Project will impact science and society in multiple ways, including:
- Domain experts have noted that *important advances* are limited by the current diversity of data formats, management and access policies, tools for visualization and use, preservation strategies, inability to accomplish multi-disciplinary multiple platform integration, and inability to easily extract needed subsets from enormous data collections. The DataSpace infrastructure will *support that diversity* and *empower researchers* to effectively utilize all relevant data resources.
- To ensure *continual availability* of scientific data to all communities, DataSpace will be highly distributed to *manage the risk of failure* either in technology (by supporting extensibility) or organization (for example, control by a small number of entities, or an inefficient monopoly caused by large-scale centralization). Technology is still evolving rapidly and innovation is occurring in many sectors and countries, so the cyberinfrastructure must be designed to leverage that situation.
- We strongly support the vision of *Open Access* to all information, while recognizing that there are sometimes necessary constraints that must be honored. We are committed to making universal, meaningful access a high priority in the expectation that it will lead to *unexpected advances* in science and engineering.
- The DataSpace Project will produce *innovation* in several areas fundamental to data archiving infrastructure: data interoperability and visualization frameworks, large-scale database and storage infrastructure, legal and policy frameworks, broad public access, and new techniques to lower the cost of data to benefit scientific endeavors and society in general.

TABLE OF CONTENTS

For font size and page formatting specifications, see GPG section II.C.

	Total No. of Pages	Page No.* (Optional)*
Cover Sheet for Proposal to the National Science Foundation		
Project Summary (not to exceed 1 page)	1	
Table of Contents	1	
Project Description (Including Results from Prior NSF Support) (not to exceed 15 pages) **(Exceed only if allowed by a specific program announcement/solicitation or if approved in advance by the appropriate NSF Assistant Director or designee)**	7	
References Cited	13	
Biographical Sketches (Not to exceed 2 pages each)	46	
Budget (Plus up to 3 pages of budget justification)	3	
Current and Pending Support	1	
Facilities, Equipment and Other Resources	0	
Special Information/Supplementary Documentation	3	
Appendix (List below.) **(Include only if allowed by a specific program announcement/ solicitation or if approved in advance by the appropriate NSF Assistant Director or designee)**		

Appendix Items:

*Proposers may select any numbering mechanism for the proposal. The entire proposal however, must be paginated. Complete both columns only if the proposal is numbered consecutively.

DataNet Preliminary Proposal: MIT DataSpace - Project Description

I. Introduction

Many aspects of society have been transformed by the Web – the large-scale, sustainable cyberinfrastructure that includes the Internet, the World Wide Web and related technologies, and the rich ecosystem of services and social practices that have grown in that environment (Amazon, Google, eBay, Yahoo, and innumerable other Web-based services). This cyberinfrastructure has transformed business and many other industries over the past decade not only because of its highly distributed, flexible architectures and protocols, but also because a legal and policy framework and a new set of social norms have emerged. So far, science has not seen the same transformative benefit from this infrastructure. The proposed DataSpace Project begins with the existing Web and other relevant technologies and relationships to model a new type of flexible, scalable infrastructure and organization that will collectively build out the scientific research data component of the cyberinfrastructure.

Different domains (and sub-domains) of scientific research require different data formats, management and access policies, tools for visualization and use, and preservation strategies. Domain experts have noted that important scientific advances are being limited by the inability to accomplish multi-disciplinary multiple platform integration and by the inability to easily extract needed data subsets from enormous data collections. The DataSpace infrastructure will be designed to support that diversity and empower any researcher to effectively utilize all the data resources across multiple domains – thereby potentially transforming the nature of scientific research. In addition, the DataSpace Project will produce innovations in several areas fundamental to data archiving to benefit science and society in general.

The Massachusetts Institute of Technology (MIT) has played a key role in the creation of many of today's core infrastructural technologies – the Internet, the World Wide Web, the Semantic Web, Kerberos, X-Windows, and others – and continues to provide leadership in envisioning, researching, prototyping, and operationalizing new infrastructural technologies. MIT is already working on many aspects of the data curation and interoperability central to the idea of cyberinfrastructure, such as:

- performing scientific research that makes extensive use of current technical infrastructure and exemplifies the opportunities and challenges faced by scientists and engineers in managing data
- significant research into data quality, interoperability and representation standards
- development of the emerging Semantic Web
- expertise on platforms for data ingest, curation, preservation and reuse
- extensive technology operations experience with network security, privacy and scalability
- development of policies and legal frameworks for digital data and archives management, both internally and cross-organization
- innovations in data visualization and navigation technology
- a strong commitment to open access to research and teaching material
- research on and experience creating sustainable business models for distributed organizations

MIT proposes a project to bring together these ongoing threads of work into a coherent program of research, development, operations and outreach that will serve to both define a model data archive cyberinfrastructure and be an exemplar for implementing the model that can scale well beyond what is currently achievable. The project will build on current expertise in data curation, interoperability and integration (e.g. the MIT Productivity from Information Technology research agenda and the DSpace open source digital archive platform) to create an exemplar for an institutionally-based data archive program that is locally sustainable, can be deployed at any research organization, is federated into a coherent whole, and will scale over time to meet the needs of many types of research-generating organizations and scientific domains.

The resulting infrastructure should enable research organizations to join a global data network of archives as easily as they can join the Web. We will separate the concerns of physical data storage and low-level preservation, data curation and functional preservation, data access and visualization across different sectors and belonging to the community of practice that best serves each piece of the problem. The infrastructure will depend on a set of practices, standards and protocols that are independent of

particular hardware or systems, distribute and replicate the data, and provide federated access to all the relevant communities of interest.

Building the science cyberinfrastructure for the twenty first century will be an enormous undertaking requiring patience, determination, and rigor. The DataSpace Project has a team that can build a complete operational system. Once the model exists we can continue to extend it for decades to come, and because it's open and enables many forms of business we can create a new ecosystem and incentive system for a science infrastructure that will be transformative and sustained.

II. Scientific Research Challenges

The DataSpace Project recognizes the central role of scientific research practices in designing relevant data archiving and future usage strategies, and has solicited the involvement of senior researchers working in two major scientific domains to initially inform its work: life sciences and energy/environment. In particular, we will collaborate with researchers working with bioscience data generated at the Broad Institute, a joint research program of MIT and Harvard University, and with researchers working on climate change data gathering and model outputs associated with the MIT Joint Program on the Science and Policy of Global Change. Both domains have identified data interoperability, aggregation, and longevity as key problems, as well as the need to bring their data into modern technological infrastructure that will lower the cost of data processing and support improved interdisciplinary data use.

- Life Sciences

In the past twenty years the life sciences – biology, biochemistry, public health, etc. – have become a dominant branch of science, generating enormous quantities of genomic, proteomic, phenotype, clinical trial, computer-generated image, and many other types of data to describe biological phenomena and their effect on public health and the environment. Researchers in this domain have identified the challenge of data sharing and integration as a high priority. The problem includes not only sharing of primary data, which in the case of unstructured data such as images poses a great challenge, but also sharing of numerical models and methods/algorithms that were used to build the models. These models, as well as ever increasingly large databases, are the most effective way of disseminating information.

- Energy/Environment Sciences

Studying the effect of economics and other predictors on global climate change is an enormously complex, data-driven undertaking involving aging, non-standard data formats that are expensive to process and cannot easily be used by other disciplines, although energy research is highly interdisciplinary. The data processing software is often designed for custom, binary formats requiring highly specialized knowledge to interpret. As the amount and complexity of this data and its associated mathematical models increase, the need for defining modern ontologies and data sharing standards and best practices has become acute.

These two domains of scientific research demonstrate a key rationale for the NSF DataNet program: they both pose many challenges related to data expression, encoding, documentation, sharing, visualizing, and preserving; as a result, it is difficult today to perform research that crosses sub-domains due to the use of disparate data. Furthermore, an even greater challenge is performing research across these two domains despite obvious benefits of doing so - the effects of global climate change on public health and genomics are important research topics that are impractical today given the difficulty of developing models across the existing data from these two domains.

III. Infrastructure Design and Research Agenda

The infrastructure that will be designed for the DataSpace Project builds on existing infrastructure and extends it to Web-scale. We will create a basic operational system in the first part of the project, while doing the research necessary to evolve that infrastructure to a next generation design that will allow for very large scale and broadly distributed network of archives.

- *DSpace archive platform*: Originally created by HP Labs and MIT in 2000-2002, DSpace is open source software designed to implement the Open Archives Information System (OAIS) reference model for long-term digital archives. It is in use by approximately 300 research institutions world wide for access to and long-term archiving of research output. It includes functionality for data deposit, management, discovery, and preservation, and assigns globally unique, persistent identifiers to archived material. It facilitates "open" access but supports controlled access and embargoes where necessary.

Prior research on the DSpace platform related to long-term data curation includes work by HP Labs on federated, replicated archives; the SPECTRa Project at the University of Cambridge on workflows for chemistry (crystallography) data deposit; work by MIT and the San Diego Supercomputer Center on a policy framework for locally-controlled, globally-enforced distributed data curation; and several projects at MIT and other institutions in the DSpace community on digital preservation strategies for a variety of data formats.

Although the DSpace platform will be a key part of the DataSpace Project's initial operational infrastructure, we anticipate major changes in the storage, curation and user interface functions to support large-scale scientific data archiving. Furthermore, a key design goal of the DataSpace Project is to achieve platform independence by definition of standards and protocols that can be implemented by any archiving systems so that a competitive market can emerge (similar to the range of options for Web servers and browsers that exists today). We will identify the set of standards and protocols, a working implementation, and an outreach plan that will encourage additional implementations in the future.

- *Data protection and security*: While MIT and the DataSpace Project are committed to the principle of open access to research data wherever possible, we acknowledge that a lot of data cannot be shared for national security reasons, or to respect the privacy of data sources (e.g. HIPAA regulations). The DataSpace infrastructure will support this requirement, building on solid, mature security protocols such as Kerberos (developed by MIT) and PGP (based on patents held by and freely available from MIT).

- *Data discovery and data semantics*: Discovery of scientific research data is one of the largest problems facing successful deployment of a global cyberinfrastructure. Unlike text, which can be easily indexed by search engines, data is not immediately findable or usable without metadata to describe its contents, structure and meaning. This metadata is often implicit in the databases and other technologies that house it. In addition to metadata, it is important to fully understand the semantics of the metadata. Knowing that a piece of data is about temperature is one level of meaning, but knowing in which way temperature is recorded (e.g. $^\circ$F, $^\circ$C, or $^\circ$K), is also important. Domain specific ontologies will be used to represent this knowledge – and, since DataSpace is intended to store data over decades, it is important that "temporal semantics" also be provided (e.g. in Russia, dates were recorded using the Julian calendar until 1918, subsequently dates used the Gregorian calendar.) For these reasons the DataSpace Project will focus on metadata and ontology creation – both automatic and human generated. This is an area of particular expertise in libraries and archives that we can leverage

- *Data quality:* For data to be reused, it must be trusted, so it it should include extensive quality information, such as provenance: the details of how, when, where, why, and by whom it was produced.

- *Data interoperability and integration*: Identifying standards for data encoding can make it easier to interoperate in a federated, distributed, interdisciplinary environment. This is the single biggest challenge, and greatest win, facing the DataSpace initiative. Storing and preserving data for the long term is only useful if that data can be interpreted in the future. Integrating research data within and across disciplines is enormously difficult today, whether to ask questions across datasets or to link data with other research outputs (e.g. the published literature). The DataSpace Project will exploit emerging Semantic Web standards to achieve this data interoperability, while acknowledging that meaningful data integration will always require a large degree of knowledge about data semantics. Our organizational model includes library/domain experts to provide this data curation service by exploiting local knowledge of data production methodologies and semantics with international data ontology standards efforts.

- *Data conversion*: Although DataSpace will promote common data standards, heterogeneous data will persist - due to its legacy heritage as well as particular needs of different domains and sub-domains of science. DataSpace will provide the means for automated generation of data conversion programs, based on MIT's Context Interchange technology, that will enable the easy conversion of data to facilitate data integration.

- *Data analysis and visualization*: Like data encoding, long-term archiving and preservation of data is only useful if the tools to analyze and visualize the data also persist over time. Designing useful data visualizations is an active area of research itself – involving statistics, human/computer interaction, cognitive psychology, and systems modeling. The absence of well-developed methods for multivariate data analysis, data visualization, and spatial modeling of dynamic systems represents three major obstacles that currently limit use of image-derived data for systems modeling.

- *Data storage*: Given the size and growth rate of scientific research data, storage for a scalable DataSpace cyberinfrastructure will require a fundamentally new architecture. Storage services need not be co-located with curation and higher-order preservation activities, thus enabling curating organizations to use 3rd party storage service providers, with or without "persistence" services providing basic bit-level data preservation. Other industries – finance, insurance, commercial engineering, etc. – also face large-scale storage needs and are creating a vibrant commercial market in this area. The DataSpace Project will initially work with two of its external partners (RENCI and HP Labs) to determine storage requirements and near-term architectural options with commercial or non-profit service providers (e.g. Internet Archive, Akamai, Iron Mountain), while also investigating new storage architectures such as those used by Amazon and Google to develop requirements for next generation storage systems.

- *Intellectual property*: Infrastructure is not simply a technical issue. The law is a significant component of the network infrastructure for science. Copyrights, privacy rights, patent rights, data rights, and more all need to be managed. Data sharing and reuse are today plagued by varying practices and legal frameworks. The DataSpace Project will collaborate with the Science Commons to develop licenses and best practices for open data sharing internationally, and promote those practices through outreach and the curation systems we deploy.

- *Policy management:* Similarly, a science cyberinfrastructure needs a tested, scalable policy framework that allows local curatorial control over data management while being recognizable and enforceable across the entire network of archives and service providers. This will be done through collaboration with the W3C, the Science Commons, and others active in this area. Data integration across heterogeneous sources requires not only semantic integration, but also identification and resolution of policy-based restrictions on dynamically-integrated data. With Science Commons policy metadata, copyright information presented in Semantic Web formats with policies can be evaluated and acted upon in a machine-assisted manner.

- *Internet domain*: Additionally, we will explore the creation of a new top-level Internet domain – .arc – that would be granted by an accreditation process (e.g. based on the Trusted Digital Repository checklist) and with certain functional/policy guarantees. This would allow both producers and consumers of data to understand an archive's data curation intent and reliability.

- *Workflows for research and archives*: Data curation is a holistic process that begins with data creation and is ongoing throughout the data's entire lifecycle, so successful data capture, including useful documentation, must be integrated into the research workflow of each domain. Today this would involve custom integration of an infinite number of custom systems, so extensive work is needed to identify or develop protocols to support simple integration of curation and preservation activities with other parts of the data lifecycle.

- *Business models for archives and distributed organizations*: The model for sustainability of the DataSpace archives network will depend on its distributed, multi-purpose nature, so that the cost is not borne by researchers alone. The institutions that support research and other industries that share many of the DataNet's infrastructure needs can also bear the costs of the infrastructure, much like the Internet.

- *International consensus*: A key requirement for the long-term sustainability of DataSpace is an institutional mechanism to evolve and gain consensus support for technical and operational standards. As the global body responsible for setting Web standards and defining the future architecture of the Web, the World Wide Web Consortium (W3C) offers expertise and an existing institution that can build consensus on DataSpace-derived standards. W3C could lead the standardization effort alone or in partnership with other standards groups in the library, publishing and archival communities.

IV. Educational and Social Outreach

Use of science and engineering data in educational settings is growing as active learning methods become more accepted. While it is out of scope for this project to work on changing the science and engineering curriculum to make more use of data, we will actively support cases where this happens by making the data citable and reusable, and documenting how data was used in teaching via MIT's OpenCourseWare (OCW) program, which publishes the materials of every MIT course as freely available content on the Web. OCW sites will include archived data and related documentation (e.g. visualization tools, problem sets using the data). We will also liaise with popular open source software course management systems (e.g. Sakai, Moodle) to insure ease of incorporating available data into coursework.

In addition, we will explore ways, in conjunction with MIT's Center for Collective Intelligence, to encourage and empower society at large to become more actively engaged in scientific investigation and the use of scientific data.

V. Scholarly Publishing

Another major new use of scientific research data is in "enhanced publications" that link published research articles with the data that supports their findings. Publishers in the several scientific domains are beginning to incorporate this functionality into their publishing platforms (e.g. BioMedCentral, PLoS, and Nature Publishing Group) and would benefit from more standardized access to the data, both in format and means of citation (e.g. persistent, Web-enabled identifiers and standards for data description and provenance). The DataSpace Project will work with the publishing community to insure the use of the data encoding and metadata standards in publishing for improved publication quality and enhanced functionality.

VI. Assessment Plan

The DataSpace Project will create both a high level Advisory Board (discussed below) as well as an active Management Board to oversee its activities. The responsibility of these boards will be to help the project develop both realistic operations plans and business models for ongoing sustainability.

VII. International Relations

The DataSpace Project team has many existing international ties that it can exploit for outreach and coordination with related efforts in other countries. Both HP Labs and the DSpace Foundation are international organizations with representation in all parts of the world, including the developing world. They have particularly strong ties to China and India, and throughout the UK, Europe and Australia. The W3C is an international organization that coordinates the standards and protocols underlying the Web. The Web Science Research Initiative is a new joint program with the University of Southampton in the UK (the source of the EPrints open source digital repository platform) that coordinates international research around all aspect of the Web including social and economic research.

Our proposed Advisory Board includes major international research organizations such as ERCIM and CODATA, as well as related cyberinfrastructure initiatives in the UK, Europe, Australia and elsewhere.

VIII. Organization

It is our belief that research-generating organizations can and should play an active role in curating their own data. Once the cost of doing so are in line with other enterprise operations (e.g. Internet connectivity, Web site management, Library and Course Management Systems) they will accept that responsibility, including ongoing financial support. Using MIT as an initial exemplar, and partnering with other major research universities (Rice and Georgia Tech, initially) to test the operational and business models we develop, we will demonstrate the technical and economic viability of a distributed, federating model of data archives.

The archive network that emerges from the DataSpace Project will be highly decentralized, flexible, and multi-modal. It will be light-weight and low-cost, to lower barriers to adoption by any organization that generates research data. A new organization will serve to coordinate these distributed activities, similar to the role of ICANN or the W3C. Since much of the DataSpace network's interoperability will build on existing infrastructure and standards, the added coordination costs of the curation activities can be kept low.

IX. DataSpace Project Staff and Partners

<u>MIT Research and Operations Team</u>

•   The Principal Investigator is Stuart Madnick, the John Norris Maguire Professor of Information Technology at the MIT Sloan School of Management and Professor of Engineering Systems at the MIT School of Engineering. Professor Madnick has had a long-term research agenda that addresses ways to integrate information systems, giving organizations a more global view of their operations. He has led a project to develop new technologies for gathering and analyzing information from many different sources, including conventional databases and the World Wide Web. He has tested these new technologies in industries such as financial services, manufacturing, logistics, counter-terrorism, and transportation.

•   The co-PIs for the DataSpace Project include Professor Hal Abelson (Class of 1922 Professor of Computer Science and Engineering in the Department of Electrical Engineering and Computer Science),

Jerry Grochow (MIT's CIO and Director of its central IT department), and MacKenzie Smith (Associate Director for Technology in the MIT Libraries). The PIs collectively reflect the combination of research and operations expertise that are required to achieve the DataSpace program goals.

- Project Manager – a dedicated Project Manager will be recruited to coordinate the activities of the team and liaise with the NSF and other DataNet Partners.

- Libraries – will provide domain experts to work directly with scientists to identify ontologies and other standards for data description and encoding; to perform ingest, description, and linking activities; to define curation policies for the archives.

- Information Services & Technology – will design and operate enterprise DataSpace infrastructure including storage, security, system support, database administration, and other core enterprise operations.

- CSAIL and Engineering Systems Division – will perform research and prototyping of new technologies required for the DataSpace infrastructure, including data visualizations and visualization frameworks, data interoperability standards and protocols, and innovative database designs.

- Sloan School of Management – will provide expertise on several important DataSpace topics, such as incentives for data sharing and re-use, data governance policies, and sustainable business models.

Science Domain Expert Partners

- *Life Sciences*: The **Broad Institute** is a research collaboration involving faculty, professional staff and students from MIT and Harvard and is governed jointly by the two universities. It supports scientists in constructing new powerful tools for genomic medicine, to make them accessible to the global scientific community, and to apply them to the understanding and treatment of disease. The institute works on a range of scientific programs (cancer, medical and population genetics, etc.) and platforms, including imaging, biological samples, chemical biology, gene sequencing and analysis, proteomics and RNAi. The Broad Institute and other life science programs at MIT are committed to open access to their data and software. The **Harvard Medical School Center for Biomedical Informatics** (CBMI) is a research center within the Harvard Medical School that promotes and facilitates collaborative activities in biomedical informatics among researchers. The CBMI emphasizes interinstitutional, interdisciplinary, and highly collaborative research and is addressing the challenges of biomedical knowledge overload, high-throughput biomedical data generation, consumer access to biomedical information, and demands for real-time information-based public health.

- *Energy/Environment*: The **MIT Joint Program on the Science and Policy of Global Change** is an interdisciplinary organization that conducts research, independent policy analysis, and public communication on issues of global environmental change. The program combines the capabilities of two other MIT research centers: the **Center for Global Change Science** (CGCS) and the **Center for Energy and Environmental Policy Research** (CEEPR). Resources of the parent centers are strengthened by links to the Marine Biological Laboratory's Ecosystems Center in Woods Hole, the MIT Climate Modeling Initiative, and other MIT environmental programs. The Program's work is focused on the integration of natural and social science aspects of the climate issue and producing analyses relevant to ongoing national and international discussions. Cooperative efforts engage the Program with leading research institutions and non-profit organizations worldwide. The Program's cornerstone is the MIT Integrated Global System Model (IGSM) of economic and environmental change. The IGSM is a comprehensive research tool for analyzing potential anthropogenic global climate change and its social and environmental consequences. The IGSM includes consideration of climate science, technical change, and economic and social science, in an interacting set of computer models. Results of the Program's research and analyses efforts are made publicly available and are communicated directly to international and national policy-making bodies, and to other investigators.

External Partners

- *HP Labs*: HP Labs is a global research group that helps to shape HP strategy and invests in fundamental science and technology in areas of interest to HP. HP Labs has identified digital data archiving as a core technology for its customers in a variety of industries, and performs research on multiple aspects of the problem. Two groups at HP Labs – the Digital Media Systems Group (that originally created the DSpace platform) and the Advanced Business Intelligence Lab – will collaborate with the DataSpace Project on federation architectures, policies for data curation, and high-throughput transactional database technologies.

- *DSpace Foundation*: In 2007 MIT and HP jointly created the DSpace Foundation, a non-profit organization to provide leadership to, and coordinate the activities of, the approximately 300 organizations world-wide using and contributing to the DSpace software and related activities, to manage the long-term development strategy, and to provide a forum for outreach and education of the DSpace community around new initiatives such as DataSpace.

- *Science Commons*: Science Commons develops free solutions for faster, more efficient Web-enabled scientific research by building a toolkit of policy contracts and technology. Science Commons targets areas where barriers to research are most common: access to research literature and materials, and increasing the utility of online data. Science Commons' mission is to accelerate the research cycle – the continuous production and reuse of knowledge that is at the heart of scientific method. Science Commons is a project of the Creative Commons whose copyright sharing licenses cover ninety million digital objects on the Web today.

- *Web Science Research Institute (WSRI) and Decentralized Information Group (DIG):* WSRI is a partnership between MIT and the University of Southampton to understand the scientific, technical and social factors that drive the growth of the Web and provide a global forum for academia, government and industry to collaborate. MIT's DIG explores technical, institutional and public policy questions necessary to advance the development of global, decentralized information environments. DIG is led by Tim Berners-Lee, inventor of the World Wide Web, and Daniel Weitzner, a leader in the Internet public policy field and Technology and Society Policy Director of the World Wide Web Consortium.

- *Renaissance Computing Institute (RENCI)*: RENCI is a collaborative venture of Duke University, North Carolina State University, the University of North Carolina at Chapel Hill and the state of North Carolina. RENCI performs research on storage and data management and is currently deploying a large-scale, high performance data storage system focused on research into data management and transfer of large-scale datasets. For the DataSpace project, it is anticipated that new paradigms for hardware and software need to be designed, developed and tested to address questions as to what new methods, management structures, and technologies will successfully manage the diversity, size, and complexity of data sets and data streams. RENCI will contribute to this DataSpace collaboration proposal with access to their research storage hardware facilities and staff expertise in these areas.

- *Rice University and Georgia Tech*: these institutions both conduct significant research in areas common to MIT, as well as using and contributing to DSpace. Library staff from these institutions will inform the DataSpace Project's requirements, test and provide feedback on the platform, and work with their local scientists (e.g. high energy physicists at Rice and plant biologists at Georgia Tech) on additional requirements and data formats than those specified by MIT's domain experts, to help insure cross-disciplinary effectiveness of the infrastructure.

X. DataSpace Project Organizational Structure

To ensure the effective management and evolution of the DataSpace project, there will be both a Management Board and Advisory Board as well as a long-lasting DataSpace Federation:

- *Management Board*: This will consist of all the PIs and Senior Researchers and will meet at least once a month to review progress and set near-term and long-term goals and plans. A key goal of the management board is to ensure good collaboration and communication between and among the domain experts and the DataSpace researchers and operations teams.

- *Advisory Board*: This will consist of a diverse ensemble of knowledgeable leaders from throughout the world and diverse industries (including corporate as well as scientific) to ensure that "best practices" are understood and adopted. The Advisory Board will meet at least twice a year and will be available for advice on an ongoing basis. Besides providing insights to the Management Board, the Advisory Board will be one of the mechanisms to achieve outreach to communities that are potential users and supporters of DataSpace. A number of notable individuals have already agreed to join the Advisory Board (listed in Bio Sketch index.)

- *DataSpace Federation*: This will initially consist of the early adopters of the DataSpace technology and will evolve, over time, to be the sponsoring/coordinating body for the sustainment of DataSpace, much like the DSpace Federation.

DataNet Preliminary Proposal: MIT DataSpace Project - **References**

[ACC*02]   D. Abadi, D. Carvey, U. Cetintemel, M. Cherniack, C. Convey-Sangdon, M. Stonebraker, N. Tatbul, S. Zdonik, **"Aurora: A New Model and Architecture for Data Stream Processing",** Proc. 2002 VLDB Conference, Hong Kong, China August 2002

[ADW01]   Ackerman, M., Darrell, T., Weitzner, D. **"Privacy in context"**. Human-Computer Interaction (16), pp. 167-176, 2001

[BBC*04]   H. Balakrishnan, M. Balazinska, D. Carvey, U. Cetintemel, M. Cherniack, C. Convey.,Eddie Galvez, J. Salz, M. Stonebraker, R. Tibbetts, S. Zdonik**, "Retrospective on Aurora,"** VLDB Journal 13(4), p370-383 (2004)

[BCH02]   D. Burdett, Q. Chen, M. Hsu, **"Conductor: An Enabler for Web Services-based Business Collaboration."** IEEE Data Engineering Bulletin 25(4): December 2002 pp. 22-26

[BCW06]   L. Kagal, T. Berners-Lee, D. Connolly, D. Weitzner, **"Using Semantic Web Technologies for Policy Management on the Web**", 21st National Conference on Artificial Intelligence (AAAI), July 16 - 20, 2006.

[BFG*97a]   S. Bressan, K. Fynn, C. Goh, S. Madnick, T. Pena, and M. Siegel, "**Overview of a Prolog Implementation of the COntext INterchange Mediator**", *Proceedings of the Fifth International Conference and Exhibition on the Practical Applications of Prolog*, London, England, April 1997.

[BGF*97b]   S. Bressan, C. Goh, K. Fynn, M. Jakobisiak, K. Hussein, H. Kon, T. Lee, S. Madnick, T. Pena, J. Qu, A. Shum, and M. Siegel, "**The Context Mediator Prototype**", *Proceedings of the ACM Conference on the Management of Data (SIGMOD '97)*, February 1997, pp. 525-527 [http://web.mit.edu/smadnick/www/papers/P057.pdf].

[BGL*00]   S. Bressan, C. Goh, N. Levina, S. Madnick, A. Shah, M. Siegel, "**Context Knowledge Representation and Reasoning in the Context Interchange System**", *Applied Intelligence: The International Journal of Artificial Intelligence, Neutral Networks, and Complex Problem-Solving Technologies*, Vol. 12, No. 2, September 2000, pp. 165-179.

[BGL*97]   S. Bressan, C.H. Goh, T. Lee, S. Madnick, and M. Siegel, "**A Procedure for Mediation of Queries to Sources in Disparate Context**", *Proceedings of the International Logic Programming Symposium*, October 1997.

[BHH*06]   T. Berners-Lee, W. Hall, J. Hendler, N. S, Daniel J. W, **"Creating a Science of the Web**", Science, 313(11), August 2006.

[BHM90]   P. Bernstein, M. Hsu, B. Mann, **"Implementing Recoverable Requests Using Queues"** Proc. 1990 ACM Intern. Conf. on Management of Data (SIGMOD), May 1990 pp. 112-122.

[BL99]   Berners-Lee, Tim, **"Weaving the Web"**, Harper, San Francisco, 1999.

[BLC*99]   Tim Berners-Lee, Dan Connolly, Ralph R. Swick "Web Architecture: Describing and Exchanging Data", W3C Note, 1999/6-7.

[BLC96]   T. Berners-Lee, D. Connolly **"Hypertext markup Language - 2.0"**, RFC1866, 1996/5.

[BLH*01]   Berners-Lee, Tim; Hendler, James and Lassila, Ora "__The Semantic Web__", Scientific American, May 2001, pp.29-37.

[BLH*06]   Tim-Berners Lee, Wendy Hall, James A. Hendler, Kieron O'Hara, Nigel Shadbolt and Daniel J. Weitzner, **"A Framework for Web Science", Foundations and Trends in Web Science"**, ISBN: 1-933019-33-6, September 2006, 144pp.

[BLH01]   Berners-Lee, Tim. and Hendler, James **"__Publishing on the Semantic Web__"**, Nature, April 26 2001 p. 1023-1025.

[BLM*94]   T. Berners-Lee, L. Masinter, M. McCahill, **"Universal Resource Locators (URL)"**, RFC1738, 1994/12.

[BM01]   S. Bhalla and S. Madnick, "**Parallel (on the fly) Reading of an Entire Database Copy**"*, International Journal of Computer Research*: Special Issue on Practical Applications of Parallel

Computing, Vol 10, No. 4, 2001, also in *Practical Parallel Computing*, (Editors: Marcin Paprzycki, Luciano Tarricone, Laurence Tianrou Yang), Nova Science Publishers, ISBN 1-59033-127-3, pp. 149-163, 2002 [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=335803].

[BM02]   S. Bhalla and S. Madnick, "**Asynchronous Reading of Consistent Copy of a Large Database**", *Advances in Computation: Theory and Practice, Vol. 12* (Practical Applications of Parallel Computing – 2002), (ISBN 1-59033-532-5), 2002.

[BM04]   S. Bhalla and S. Madnick, "**Asynchronous Backup and Initialization of a Database Server for Replicated Database Systems**", *Journal of Supercomputing* (Kluwer Academic Publishers), Vol. 27, No 1 (Jan. 2004), pp. 69-89.
[http://papers.ssrn.com/sol3/papers.cfm?abstract_id=391840].

[BMG04]   S. Bhansali, S. Madnick and B. Grosof, "**Towards Ontological Context Mediation for Semantic Web Database Integration: Translating COIN Ontologies into OWL**", *Proceedings of the 3rd Intl. Semantic Web Conference* (ISWC-2004), Nov. 2004.

[BMW03] D. Ballow, S. Madnick, and R. Wang, "**Assuring Information Quality,**" *Journal of Management Information Systems, Vol. 20*, No. 3, Winter 2003-04, pp. 9-11.

[BRJ99]   M. Babiker, J.M. Reilly, H.D. Jacoby, **"The Kyoto Protocol and Developing Countries"**, *Energy Policy* 28:525-536, 1999

[BSG*00]   Bresson, S., C. Goh, N. Levina, A. Shah, M. Siegel**, "Context Knowledge Representation and Reasoning in the Context Interchange System,",** The International Journal of Artificial Intelligence, Neutral Networks, and Complex Problem-Solving Technologies, Volume 12, Number 2, September 2000, pp. 165-180, [SWP #4133, CISL #00-04]

[BTG*04]   Brunet J-P, Tamayo P, Golub T, Mesirov JP. "**Metagenes and Molecular Pattern Discovery Using Matrix Factorization**", Proc. Natl. Acad. Sci. USA. 2004; 101(12):4164-4169

[BTS04]   Bass, M., Tansley, R., Smith, M., **"The DSpace Institutional Repository System: Status, Roadmap, Research, Governance, and Community Building",** Proceedings of the International Conference on Digital Libraries, February 2004

[BW95]   J. Berman, D. 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

[BW97]   J. Berman, D.Weitzner, **"Directing Policy-Making Around the Net's Metaphor**", Communications of the ACM v.40 February 1997, pp.83-84

[CBM*99]   C. Goh, Cheng, S. Bresson, Stuart Madnick, and Michael Siege, **"Context Interchange: New Features and Formalisms for the Intelligent Integration of Information",** Transactions on Information Systems, July 1999

[CEE*07]   U.S. CCSP, 2007: "**U.S. Climate Change Science Program, Synthesis and Assessment Product 2.1A: Scenarios of Greenhouse Gas Emissions and Atmospheric Concentrations"** (L. Clark, J. Edmonds, H. Jacoby, H. Pitcher, J. Reilly, R. Richels), in press, 2007 (http://www.climatescience.gov/Library/sap/sap2-1/default.php).

[CEG*06]   Nazli Choucri, Christi Electris, Daniel Goldsmith, Dinsha Mistree, Stuart Madnick, J. Bradley Morrison, Michael Siegel, Margaret Sweitzer-Hamilton, **"Understanding & Modeling State Stability: Exploiting System Dynamics",** System Dynamics Conference, Cambridge , MA January 2006

[CHD00]   Q. Chen, M. Hsu, U. Dayal, "**A Data Warehouse/OLAP Framework for Scalable Telecommunication Tandem Traffic Analysis"**, *Proc. of 16th IEEE International Conference on Data Engineering (ICDE),* March 2000, pp. 201-210

[CHD01]   Q. Chen, M. Hsu, **"Inter-Enterprise Collaborative Business Process Management,"** *Proc. International Conference on Data Engineering (ICDE),* 2001.pp 253-260

[CKM*03]   W. Chu, C. Knoblock, S. Madnick and S. Ram, "**Information and Knowledge Sharing – Successes and Failures**,", *Proceedings of the IASTED International Conference on Information and Knowledge Sharing*, Scottsdale, Arizona, November 2003, in ISBN 0-88986-396-2, Acta Press, Anaheim, CA, pp. 3-5.

[CMM*04]   N. Chourci, S. Madnick, A. Moulton, M. Siegel, and H. Zhu, "**Information Integration for Counter Terrorism Activities: The Requirement for Context Mediation,"**, *Proceedings of the 2004 IEEE Aerospace Conference*, Big Sky, Montana, March 6-13, 2004.

[CP05]   Chen, Y.-H., R.G. Prinn, **"Atmospheric modeling of high-frequency methane observations: Importance of interannually varying transport",** J. Geophys. Res., 110, D10303,

doi: 10.1029/2004JD005542, 2005

[CP06]   Chen, Y.-H., R.G. Prinn, **"Estimation of atmospheric methane emissions between 1996 and 2001 using a 3D global chemical transport model"** J. Geophys, Res., 111, D10307, doi: 10.1029/2005JD006058, 2006

[D88]   P. Dreher **"Using Microcomputers for Teaching Lattice Gauge Theory Computational Techniques"**, Computers in Physics, 2(6), 49, (1988).

[D93]   P. Dreher, **"Study of Vortex Excitations in 3 Dimensional Abelian Higgs Models Using Scientific Visualization",** Proceedings Division of Computational Physics, Sante Fe (1993)

[DAC*02]   P. Dreher, W. Akers, J. Chen, Y. Chen, C. Watson, **"Physics development of Web-based Tools for Use in Hardware Clusters Doing Lattice Physics"** XXIX International Lattice Field Theory Conference, Berlin (2001) **N**uc. Phys. B (Proc) 106-107, 1040 (2002).

[DGH*95]   A. Daruwala, C. Goh, S. Hofmeister, K. Hussein, S. Madnick, and M. Siegel, "**The Context Interchange Network Prototype**", *Proceedings of the Sixth IFIP TC-2 Working Conference on Data Semantics (DS-6)*, May 1995.

[DGH*96]   A. Darawala, C. Goh, S. Hofmeister, S. Madnick, and M. Siegel, "**Context Interchange Network Prototype**"in *Database Applications Semantics*, Chapman & hall (London, UK), October 1996.

[DGMS75]   J.J. Donovan, L.M. Gutentag, S. Madnick, and G.N. Smith, "**An Application of a Generalized Management Information System to Energy Policy and Decision Making: The User's View**", *Proceedings AFIPS 1975 National Computer Conference*, Vol. 44, May 1975.

[DHL01]   U. Dayal, M.Hsu, R. Ladin, **"Business Process Coordination: State of the Art, Trends, and Open Issues"**, 27th International Conference on Very Large Data Bases (VLDB), September 2001. pp.3-13.

[DHL91]   U. Dayal, M. Hsu, R. Ladin, **"A Transaction Model for Long-Running Activities"**, Proc. of the 17th Int. Conf. on Very Large Databases (VLDB), Sept., 1991, pp 113-122

[E01a]   J. Erickson, **"A Digital Object Approach to Interoperable Rights Management: Fine-grained Policy Enforcement Enabled by a Digital Object Infrastructure."** *D-Lib Magazine* June, 2001).

[E01b]   J. Erickson, **"Information Objects and Rights Management: A Mediation-based Approach to DRM Interoperability"** D-Lib Magazine April, 2001.

[E03]   J. Erickson, **"Fair Use DRM and Trusted Computing"** *Communications of the ACM,* Vol. 46, Issue 4 (Apr 2003).

[E03]   J. Erickson, **"Digital Object Identifier"** *McGraw-Hill Encyclopedia of Science & Technology*, 9th edition (Jan 2003).

[E06]   J. Erickson **"Handle Records, Rights and Long Tail Economies."** *D-Lib Magazine,* Vol.12 No.9 (Sep 2006).

[E07]   J. Erickson, J. Rutherford; **"pf-dspace: standards-based peer-to-peer federation of DSpace repositories."** Open Repositories 2007, San Antonio, TX (Jan 2007)

[E08]   J. Erickson **"The Future of the Institutional Repository: Making it Personal."** Submission to Open Repositories 2008, Southampton, UK.

[EM04] J. Erickson, D. Mulligan **"Technical and Legal Dangers of Code-based Fair Use Enforcement"** *Proceedings of the IEEE*, Special Issue on Enabling Technologies for Digital Rights Management (Jun 2004).

[EVP*04]   J. Erickson, H. Van de Sompel, S.Payette, C. Lagoze, S.Warner) **"Rethinking Scholarly Communication: Building the System that Scholars Deserve."** D-Lib Magazine (Sep 2004).

[FGF*99]   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.

[FGM07]   A. Firat, B. Grosof and S. Madnick, "**Contextual Alignment of Ontologies in the eCoin Semantic Interoperability Framework**", *Information Technology and Management Journal*, Vol. 8, No. 1, Springer US, March 2007, pp. 47-63.

[FLMC01]   W. Fan, H. Lu, S. Madnick and D. Cheung, "**Discovering and Reconciling Value Conflicts for Numerical Data Integration**,", *Information Systems Journal*, Vol 26, No 8, pp. 635-656, November 2001 [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=335580].

[FM00]   A. Firat and S. Madnick, "**The Caméléon Approach to the Interoperability of Web**

Sources and Traditional Relational Databases,” *Proceedings of the Workshop on Information Technology and Systems*, December 2000.

[FMG02a]   A. Firat, S. Madnick and B. Grosof, “**Financial Information Integration In the Presence of Equational Ontological Conflicts**”, *Proceedings of the Workshop on Information Technology and Systems* (WITS), Barcelona, Spain, December 14-15, 2002, pp. 211-216 [BEST PAPER AWARD] [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=376840].

 [FMG02b]   A. Firat, S. Madnick and B. Grosof, “**Knowledge Integration to Overcome Ontological Heterogeneity: Challenges from Financial Information Systems**” *Proceedings of the International Conference on Information Systems* (ICIS), Barcelona, Spain, December 15-18, 2002, pp. 183-194 [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=336205].

[FMG04]   A. Firat, S. Madnick and B. Grosof, “**Contextual Alignment of Ontologies for Semantic Interoperability**”, *Proceedings of the Workshop on Information Technology and Systems* (WITS), Washington, DC, December 11-12, 2004, pp. 200-205 [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=612472], Nominated for Best Paper Award.

[FMG06]   A. Firat, S. Madnick, and B. Grosof, “**Contextual Alignment of Ontologies in the eCoin Semantic Interoperability Framework**”, *Information Technology and Management Journal*, Springer US. [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=874097]

[FMM05]   A. Firat, S. Madnick and F. Manola, “**Multi-dimensional Ontology Views via Contexts in the ECOIN Semantic Interoperability Framework**”, *Proceedings of the International Workshop on Contexts and Ontologies: Theory, Practice, and Applications*, Pittsburgh, Pennsylvania, July 9, 2005 [http://ssrn.com/abstract=729383].

[FMS00]   A. Firat, S. Madnick, and M. Siegel, “**The Caméléon Web Wrapper Engine**”, *Proceedings of the VLDB2000 Workshop on Technologies for E-Services*, September 14-15, 2000.

[FMY*05]   A. Firat, S. Madnick, N.A. Yahaya, C.W. Kuan, and S. Bressan, “**Information Aggregation using the Caméléon# Web Wrapper**”, *Proceedings of the 6th International Conference on Electronic Commerce and Web Technologies* (EC-Web 2005),” Copenhagen, Denmark, August 23 - August 26, 2005, also published in *Springer Lecture Notes in Computer Science* (LNCS) 3590, K. Bauknecht et al (Eds.), pp.76-86, Springer-Verlag Berlin, 2005 [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=771492].

[FSJ02]   C. Forest, P. Stone and H. Jacoby (2002). “**How to Think About Human Influence on Climate”,** *Forum for Applied Research and Public Policy* 17(2): 47-51, 2002

[G91]   J. Grochow, “**SAA: A Guide to Implementing IBM's Systems Application Architecture”,** Yourdon Press/Prentice Hall, 1991

[G96]   J. Grochow, “**Information Overload! Creating Value with the New Information Systems Technology”,** Yourdon Press/Prentice Hall. 1996

[GBMS99]   C.H. Goh, S. Bressan, S. Madnick, M. Siegel, “**Context Interchange: New Features and Formalisms for the Intelligent Integration of Information**”, *ACM Transactions on Office Information Systems*, July 1999. [http://web.mit.edu/smadnick/www/papers/J043.pdf]

[GCM*95]   Goh, Cheng, S. Madnick, M. Siegel**, “Context Interchange: Overcoming the Challenges of Large-Scale Interoperable Database Systems in a Dynamic Environment”,** In Proceedings of the Third International Conference on Information and Knowledge Management, Gaithersburg, MD

[GM90]   A. Gupta and S. Madnick, “**A Taxonomy for Classifying Commercial Approaches to Information Integration in Heterogeneous Environment**”, *IEEE Data Engineering*, June 1990.

[GMS94]   C. Goh, S. Madnick and M. Siegel, “**Context Interchange: Overcoming the Challenges of Large-Scale Interoperable Database Systems in a Dynamic Environment**,” *Proceedings of the International Conference on Information and Knowledge Management* (CIKM '94), November 1994, pp. 337-346 [http://web.mit.edu/smadnick/www/papers/P050.pdf].

[GST*99]   Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, **Mesirov JP**, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. “**Molecular classification of cancer: class discovery and class prediction by gene expression monitoring”,** Science. 1999 Oct 15; 286(5439):531-7

[HBL*02]   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.

[HMS02a]   M. Hansen, S. Madnick and M. Siegel, "**Data Integration using Web Services**", *Proceedings of the Second International Workshop on Data Integration over the Web* (DIWeb'02, Toronto, Canada), May 2002 [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=376822], also published in *Lecture Notes in Computer Science*, 2590, Springer-Verlag, New York, 2003 [ISBN 3-540-00736-9], pp. 165-182.

[HMS02b]   M. Hansen, S. Madnick and M. Siegel, "**Process Aggregation using Web Services**", *Proceedings of the Workshop on Web Services, e-Business, and the Semantic Web: Foundations, Models, Architecture, Engineering and Applications*, (WseBT'02, Toronto, Canada), May 2002 [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=365000], also published in *Lecture Notes in Computer Science*, Springer-Verlag, New York.

[HS91]   M. Hsu, A. Silberschatz, **"Unilateral Commit: A New Paradigm for Reliable Distributed Transaction Processing,"** Proc. 7th Intern. Conf. on Data Engineering (ICDE), April, 1991 pp 286-293

[HX07]   M. Hsu, Y. Xiong, **"Building a Scalable Web Query System"**, *Proc., 2007 Databases in Networked Information Systems (DNIS),* October 2007; in *Springer Lecture Notes in Computer Science #4777*, pp. 322-328 2007

[J04]   H. Jacoby, "**Modeling Human-Climate Interaction**, S. Sparks, C. Hawksworth (eds.), **The State of the Planet: Frontiers and Challenges in Geophysics",** American Geophysical Union, Washington, D.C. 2004

[J04]   Jacoby, H., "**Toward a Framework for Climate Benefits Estimation**", in J. Corfee-Morlot and S. Agrawala (eds.), "*The Benefits of Climate Change Policies: Analytical and Framework Issue"s,* OECD, Paris, 2004

[KHD*02]   D. Keim, M. Hao, U. Dayal, M Hsu, **"Pixel bar charts: a visualization technique for very large multi-attribute data sets"** *Information Visualization 1(1):* 20-34 2002

[KM04]   M. Kimpton, G. Mohr et al**, "Introduction to Heritrix: a web archival crawler**", proceedings for 4th international web archiving workshop, 2004

[KMS95]   H. Kon, S. Madnick, and M. Siegel, "**Good Answers from Bad Data: A Data Management Strategy**", *Proceedings of the 1995 Workshop on Information Technology and Systems*, December 1995.

[KU06]   M. Kimpton, J. Ubois, "**Web Archiving"**, 201-212, 2006

[KW94]   M. Kapor, D. Weitzner, **"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. 1994

[LCM*99]   T. Lee, M. Chams, S. Madnick, R. Nado, and M. Siegel, "**Information Integration with Attribution Support for Corporate Profiles**", *Proceedings of the International Conference on Information and Knowledge Management*, November 1999, pp. 423-429 [http://web.mit.edu/smadnick/www/papers/P067.pdf].

[LFG*98]   H. Lu, W. Fan, C.H. Goh, D.W. Cheung, S. Madnick, "**Discovering and Reconciling Semantic Conflicts: A Data Mining Perspective**", in *Data Mining and Reverse Engineering: Searching for Semantics*, Chapman & Hall, 1998 [SWP #3957, CISL #97-04].

[LM92]   Y. Lee and S. Madnick, "**Integrating Territories: Information Systems Integration and Territorial Rationality**", *Proceedings of the ACM SIGCPR* Conference (Cincinnati, Ohio), April 1992, pp. 221-231 [http://web.mit.edu/smadnick/www/papers/P046.pdf].

[LM96]   J. Lee and S. Madnick, "**Resolving Context Heterogeneity: Applying Insights from Semantics and Ontology**", *Proceedings of the Workshop on Information Technology & Systems*, December 1996.

[LMB98]   T. Lee, S. Madnick, and S. Bressan, "**Source Attribution for Querying Against Semi-Structure Documents**", *Proceedings of the ACM Workshop on Web Information and Data Management* (WIDM'98), Washington, DC, November 6, 1998, pp. 33-39.

[LMC02]   H. Lu, S. Madnick, and D.W. Cheung, "**DIRECT: A System for Mining Data Value Conversion Rules from Disparate Sources**", *Decision Support Systems,* 2002, pp. 19-39. [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=377900].

[LMS96a]   J. Lee, S. Madnick, and M. Siegel, "**Conceptualizing Semantic Interoperability: A Perspective from the Knowledge Level**", *International Journal of Cooperative Information Systems* [Special Issue of Formal Methods in Cooperative Information Systems], Vol. 5, No. 4, December 1996.

[LMS96b]   J. Lee, S. Madnick, and M. Siegel, "**The Context Interchange Approach to Large Scale Data Integration: Benefits and Challenges**", *Proceedings of the International Conference on Data Engineering*, August 1996.

[LRL*]   Liefeld T, Reich M, Lerner J, Gould J, Zhang P, Tamayo P, Mesirov JP. ,"**GeneCruiser: A web service for the annotation of microarray data**," (in press, Bioinformatics).

[LWG*98]   H. Lu, W. Fan, C.H. Goh, S. Madnick, D.W. Cheung, "**Discovering and Reconciling Semantic Conflicts: A Data Mining Perspective**", *Proceedings of the IFIP 2.6 Conference on Database Semantics*, October 1997; also in *Data Mining and Reverse Engineering: Searching for Semantics*, Chapman & Hall, 1998.

[M*01a]   J.P. Mesirov, "**International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome**", Nature. 2001 Feb 15; 409(6822):860-92.

[M*01b]   J.P. Mesirov, "**Erratum in: Nature 2001**," Aug 2;412(6846):565. Nature 2001 Jun 7;411(6838):720

[M*03a]   T. Malone, W. Crowston, K. G., & Herman, G. (Eds.), "**Organizing Business Knowledge: The MIT Process Handbook**," Cambridge, MA: MIT Press, 2003.

[M*03b]   T. Malone, T. W. Laubacher, R. J., & Scott Morton, M. S. (Eds.), "**Inventing the Organizations of the 21st Century**", Cambridge, MA, MIT Press, 2003

[M*04a] Malone, T. W., "**The Future of Work: How the New Order of Business Will Shape Your Organization, Your Management Style, and Your Life**". Boston, MA: Harvard Business School Press, 2004 (Translated into Japanese, Chinese, Spanish, Korean, Portuguese, and Russian).

[M*04b]   Malone, T. W. "**Bringing the market inside**", *Harvard Business Review*, April 2004, *82* (4), 106-114.

[Mad01]   S. Madnick, "**The Misguided Silver Bullet: What XML will and will NOT do to help Information Integration**," *Proceedings of the Third International Conference on Information Integration and Web-based Applications and Services* (IIWAS2001; Linz, Austria), published by Osterreichische Computer Gesellschaft (ISBN 3-85403-157-2), September 2001, pp. 61-72 [http://papers.ssrn.com/abstract=281823].

[Mad03]   S. Madnick, "**Oh, So That is What you Meant! The Interplay of Data Quality and Data Semantics**," *Proceedings of the 22nd International Conference on Conceptual Modeling* (ER'03), Chicago, October 2003; in Conceptual Modeling – ER 2003, (ISBN 3-540-20299-4, Springer-Verlag at http://www.springerlink.com/index/JF29Q65B56REB0E2), 2003, pp. 3-13 [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=436805].

[Mad87]   S. Madnick, "**Perspectives on the Effective Use, Planning, and Impact of Information Technology**," *The Strategic Use of Information Technology*, Oxford University Press, 1987 (pp. 3-14).

[Mad95a]   S. Madnick, "**From VLDB to VMLDB (Very MANY Large Data Bases): Dealing with Large-Scale Semantic Heterogeneity**", *Proceedings of the 21st VLDB Conference* (Zurich, Switzerland), September 1995, pp. 11-16.

[Mad95b]   S. Madnick, "**Integrating Information from Global Systems: Dealing with the "On- and Off-Ramps of the Information SuperHighway**", *Journal of Organizational Computing*, Vol. 5, No. 2, 1995, pp. 69-82.

[Mad96]   S. Madnick, "**Are We Moving Toward an Information SuperHighway or a Tower of Babel? The Challenge of Large-Scale Semantic Heterogeneity**", *Proceedings of the IEEE International Conference on Data Engineering*, February 1996, pp. 2-8 [also reprinted in *21st* (a Web 'Zine at http://www.vxm.com/), Vol. 1, No. 6, April 1996, pp. Speed 1-10. [http://web.mit.edu/smadnick/www/papers/P054.pdf]

[Mad99]   S. Madnick, "**Metadata Jones and the Tower of Babel: The Challenge of Large-Scale Heterogeneity**," *Proceedings of the IEEE Meta-data Conference*, April 1999.

[MBMS98]   A. Moulton, S. Bressan, S. Madnick, and M. Siegel, "**Using an Active Conceptual Model for Mediating Analytic Information Exchange in the Fixed Income Securities Industry**", April, 1998, *Proceedings of the 17th International Conference on Conceptual Modeling* (ER'98), Singapore, November 1998.

[MC94] Malone, T. W. & Crowston, K. "**The interdisciplinary study of coordination**", *ACM Computing Surveys*, 1994 (March), 26 (1), 87-119

[MCL*98]   T.W. Malone, K. Crowston, J. Lee, B. Pentland B, C. Dellarocas, **"Computer system for displaying representations of processes",** US Patent No. 5,819,270 (October 6, 1998).

[MCL*99]   T. W. Malone, K. Crowston, J. Lee, B. Pentland, C. Dellarocas, G. Wyner, J. Quimby, C.S. Osborn, A. Bernstein, G. Herman, M. Klein, E. O'Donnell, **"Tools for inventing organizations: Toward a handbook of organizational processes"**, *Management Science*, 1999, 45, 3 (March), 425-443. (Reprinted in several books.)

[MHS02]   Madnick, S., M. Hansen, M. Siegel, **"Data Integration using Web Services", Proceedings of the Second International Workshop on Data Integration over the Web**", (DIWeb'02, Toronto, Canada), May 2002 [CISL #2002-14], also to be published in Lecture Notes in Computer Science, Springev-Varlag, New York

[ML98]   T. W. Malone, R. J. Laubacher, **"The Dawn of the E-lance Economy"**, *Harvard Business Review*, Sep. – Oct. 1998, 76 (5), 144-152. (Original and revised version each reprinted in several books.)

[MLE*03]   Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E, Houstis N, Daly MJ, Patterson N, Mesirov JP, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, Groop LC. "**PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes**", Nat Genet. 2003 Jul; 34(3):267-73

[MMS01]   A. Moulton, S. Madnick and M. Siegel, "**Cross-Organizational Data Quality and Semantic Integrity: Learning and Reasoning about Data Semantics with Context Interchange Mediation**", *Proceedings of the Americans Conference on Information Systems* (AMCIS, Boston), August 2001 [http://papers.ssrn.com/abstract=289992].

 [MMS01]   A. Moulton, S. Madnick and M. Siegel, "**Knowledge Representation Architecture for Context Interchange Mediation: Fixed Income Securities Investment Examples**", *Proceedings of the First International Workshop on Electronic Business Hubs: XML, Metadata, Ontologies, and Business Knowledge on the Web* (WEBH; Munich, Germany), published by IEEE (ISBN 0-7695-1230-5, ISSN 1529-4188), September 2001, pp. 50-54 [http://papers.ssrn.com/abstract=289996; http://web.mit.edu/smadnick/www/papers/P079.pdf].

[MMS01]   A. Moulton, S. Madnick, and M. Siegel, "**Cross-Organizational Data Quality and Semantic Integrity: Learning and Reasoning about Data Semantics with Context Interchange Mediation**", *Proceedings of the Americans Conference on Information Systems* (AMCIS, Boston), August 2001.

[MMS01]   A. Moulton, S. Madnick, M. Siegel, **"Cross-Organizational Data Quality and Semantic Integrity: Learning and Reasoning about Data Semantics with Context Interchange Mediation"**, Proceedings of the Americans Conference on Information Systems (AMCIS, Boston), August 2001 [SWP #4167, CISL #01-04]

[MMS01]   Moulton, Allen, S. Madnick, M. Siegel , **"Knowledge Representation Architecture for Context Interchange Mediation: Fixed Income Securities Investment Examples",** Proceedings of the First International Workshop on Electronic Business Hubs: XML, Metadata, Ontologies, and Business Knowledge on the Web (WEBH; Munich, Germany), published by IEEE (ISBN 0-7695-1230-5, ISSN 1529-4188), September 2001, pp. 50-54 [SWP #4169, CISL #01-05]

 [MMS02a]   A. Moulton, S. Madnick and M. Siegel, "**Context Interchange Mediation for Semantic Interoperability and Dynamic Integration of Autonomous Information Sources in the Fixed Income Securities Industry**", *Proceedings of the Workshop on Information Technology and Systems* (WITS), Barcelona, Spain, December 14-15, 2002, pp. 61-66 [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=376863].

[MMS02b]   A. Moulton, S. Madnick and M. Siegel, "**Semantic Interoperability in the Securities Industry: Context Interchange Mediation of Semantic Differences in Enumerated Data Types**",

Proceedings of the Second International Workshop on Electronic Business Hubs: XML, Metadata, Ontologies, and Business Knowledge on the Web, Aix En Provence, France, September 6, 2002, pp. 883-888. [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=365021; http://web.mit.edu/smadnick/www/papers/P084.pdf].

[MMS03] A. Moulton, S. Madnick and M. Siegel, "**Semantic Interoperability in the Fixed Income Securities Industry: A Knowledge Representation Architecture for Dynamic Integration of Web-Based Information**", *Proceedings of the Thirty-sixth Annual Hawaii International Conference on System Sciences*, Hawaii, January 6 - 9, 2003 [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=376861; http://web.mit.edu/smadnick/www/papers/P092.pdf] [Nominated for Best Paper Award]

[MMS98] A. Moulton, S. Madnick and M. Siegel, "**Context Interchange on Wall Street**", Proceedings of the International Conference on Cooperative Information Systems (CoopIS'98), September 1998, pp. 271-279 [http://web.mit.edu/smadnick/www/papers/P063.pdf].

[MOW89] S. Madnick, C. Osborn and Y.R. Wang, "**Motivating Strategic Alliance for Composite Information Systems: The Case of a Major Regional Hospital**", *Journal of Management Information Science*, Vol. 6, No. 3, Winter 1989/90, pp. 99-117. [also appeared in *Proceedings of the 1989 Hawaii International Conference on System Sciences (HICSS)*, January 1989 -- Nominated for Best Paper at HICSS.] http://web.mit.edu/smadnick/www/papers/J026.pdf]

[MS02] S. Madnick and M. Siegel, "**Seizing the Opportunity: Exploiting Web Aggregation**", *MISQ Executive, Vol 1*, No. 1, March 2002, pp. 35-46. [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=303827].

[MS02] Madnick, S., M. Siegel, "**Seizing the Opportunity: Exploiting Web Aggregation**", MISQ Executive, Vol 1, No. 1, March 2002, pp. 35-46. [SWP #4351, CeB #144, CISL #01 13]

[MS04] Zhu,H., S. Madnick, M. Siegel, "**Effective Data Integration in the Presence of Temporal Semantic Conflicts**", Conference on Temporal Representation and Reasoning, July 2004

[MS07] R. Moore, M. Smith, "**Automated Validation of Trusted Digital Repository Assessment Criteria**", Journal of Digital Information 8(2), 2007.

[MS89a] S. Madnick and M. Siegel, "**Maintaining Valid Schema Integration in Evolving Heterogeneous Database Systems**", *IEEE Office Knowledge Engineering*, Vol. 3, No. 2, August 1989, pp. 9-16.

[MS89b] S. Madnick and M. Siegel, "**Schema Integration Using Metadata**", *Proceedings of the 1989 NSF Workshop on Heterogeneous Databases* (Evanston, IL), December 1989.

[MS91a] S. Madnick and M. Siegel, "**A Metadata Approach to Resolving Semantic Conflicts**", *Proceedings of the International Conference on Very Large Data Bases,* (Barcelona, Spain), September 1991, pp.133-145.

[MS91b] S. Madnick and M. Siegel, "**Context Interchange: Sharing the Meaning of Data**" (with), *SIGMOD Record*, Volume 20(4), December 1991, pp. 77-79 [http://web.mit.edu/smadnick/www/papers/J036.pdf].

[MSG91] S. Madnick, M. Siegel and A. Gupta, "**Composite Information Systems: Technologies for Resolving Semantic Heterogenities**", *Proceedings of the Workshop on Information Technology and Systems* (Cambridge, Mass.), December 1991.

[MSS01] A. Moulton, S. Madnick and M. Siegel, "**Knowledge Representation Architecture for Context Interchange Mediation: Fixed Income Securities Investment Examples**", *Proceedings of the First International Workshop on Electronic Business Hubs:* XML, Metadata, Ontologies, and Business Knowledge on the Web (WEBH; Munich, Germany), published by IEEE (ISBN 0-7695-1230-5, ISSN 1529-4188), September 2001, pp. 50-54.

[MTM*03] Monti S, Tamayo P, Mesirov JP, Golub T., "**Consensus Clustering. A resampling-based method for class discovery and visualization of gene-expression microarray data**," Functional Genomics special issue, Machine Learning. 2003;52:91-118

[MW88a] S. Madnick and R. Wang "**Integrating Disparate Databases for Composite Answers**", *Proceedings of the Hawaii International Conference on System Sciences,* Vol. II: Software, January 1988, pp. 583-592. [http://web.mit.edu/smadnick/www/papers/P029.pdf]

[MW88b] S. Madnick and R. Wang, "**A Framework of Composite Information Systems for Strategic Advantage**", *Proceedings of the Hawaii International Conference on System Sciences,*

Vol. III: *Decision Support and Knowledge Based Systems*, January 1988, pp. 35-43. [http://web.mit.edu/smadnick/www/papers/P030.pdf]

[MW89a] S. Madnick and R. Wang, "**CISL: Composing Answers from Disparate Information Systems**", *Proceedings of the 1989 NSF Workshop on Heterogeneous Databases* (Evanston, IL), December 1989.

[MW89b] S. Madnick and R. Wang, "**The Inter-Database Instance Identification Problem in Integrating Autonomous Systems**", *Proceedings of the Fifth International Data Engineering Conference* (Los Angeles, CA), February 1989, pp. 46-55 [http://web.mit.edu/smadnick/www/papers/P035.pdf].

[MW89c] S. Madnick and Y.R. Wang, "**Facilitating Connectivity in Composite Information Systems**", *ACM Database*, Vol. 20, No. 3, Fall 1989. [SWP # 1975-88; http://web.mit.edu/smadnick/www/papers/J029.pdf]

[MW90a] S. Madnick and R. Wang, "**A Polygen Model for Developing Heterogeneous Database Systems with Source Tagging Capabilities**", *Proceedings of the ICIS Conference* (Copenhagen, Denmark), December 1990.

[MW90b] S. Madnick and R. Wang, "**A Polygen Model for Heterogeneous Database Systems: The Source Tagging Perspective**", *Proceedings of the VLDB Conference* (Brisbane, Australia), August 1990, pp. 519-538.

[MW90c] S. Madnick and R. Wang, "**A Source Tagging Theory for Heterogeneous Database Systems**", *Proceedings of the ICIS Conference* (Copenhagen, Denmark), December 1990

[MWC*04] S. Madnick, R. Wang, K. Chettayar, F. Dravis, J. Funk, R. Katz-Haas, C. Lee, Y. Lee, X. Xian, S. Bhansali, "**Exemplifying Business Opportunities for Improving Data Quality through Corporate Household Research**", in *Advances in Management Information Systems* (AMIS) Monograph Series, Special Volume on Information Quality, 2004. [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=529723 .]

[MWDC01] S. Madnick, R. Wang, F. Dravis, and X. Chen, "**Improving the Quality of Corporate Household Data: Current Practices and Research Directions**", *Proceedings of the Sixth International Conference on Information Quality* (IQ2001, Cambridge), November 2001, pp. 92-104 [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=365180].

[MWX04] S. Madnick, R.Wang and X. Xian, "**The Design and Implementation of a Corporate Householding Knowledge Processor to Improve Data Quality**", *Journal of Management Information Systems, Vol. 20*, No. 3, Winter 2003-04, pp. 41-69. [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=461001].

[MYB878] T.W. Malone, J. Yates, R.I. Benjamin, **"Electronic markets and electronic hierarchies"**, *Communications of the ACM*, 1987, 30, 484-497. (Reprinted in two books.)

[MZ06] S. Madnick and H. Zhu, "**Improving Data Quality Through Effective Use of Data Semantics**," *Data & Knowledge Engineering*, Vol. 59, 2006, pp. 460-476. [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=825650]

[NBD*04] J. Negele, R.C. Brower, P. Dreher, R. Edwards, G. Fleming, Ph. Hagler, Th. Lippert, A.V.Pochinsky, D.B. Renner, D. Richards, K. Schilling, W. Schroers "**Insight into nucleon structure from generalized parton distributions**", Nuc. Phys. B129-130 (Proc) (2004) 910.

[NBD*04] J. W. Negele, R.C. Brower, P.Dreher, R. Edwards, G. Fleming, Ph. Hagler, U.M. Heller, TH. Lippert, A.V. Pochinsky, D.B. Renner, D. Richards, K.Schilling, W. Schoers Nuc. "**Insight into nucleon structure from lattice calculations of moments of parton and generalized parton distributions**", Phys. B128 (Proc) (2004) 170.

[OM01] G. Olson, M, Malone, T. W., and Smith, J. B. (Eds.), **"Coordination Theory and Collaboration Technology"**, Mahwah, NJ, Erlbaum, 2001

[OWA*98] Chris Olston, Allison Woodruff, Alexander Aiken, Michael Chu, Vuk Ercegovac, Mark Lin, Mybrid Spalding, Michael Stonebraker, **"DataSplash,"** SIGMOD 1998, p550-552

[P00] Prinn, R.G., **"Measurement equation for trace chemicals in fluids and solution of its inverse, in Inverse Methods in Global Biogeochemical Cycles**", ed. P. Kasibhatla et al., Geophysical Monographs, No. 114, American Geophysical Union, pgs. 3-18, 2000

[P04]   Prinn, R.G., **"Non-CO2 Greenhouse Gases, in The Global Carbon Cycle"**, ed. C. Field
    and M. Raupach, Island Press, Washington, D.C., pgs. 205-216, 2004
[PAB*03]   A. Pochinsky, W. Akers, R. Brower, J. Chen, P. Dreher, R. Edwards, S. Gottlieb, D.
    Holmgren, P. Mackenzie, J. Negele, D. Richards, J. Simone and W. Watson **"Large scale
    commodity clusters for lattice QCD"** Nuc. Phys. (Proc) B119 1044 (2003)
[PD05]   Prinn, R. and S. Dorling, "**Environmental Management**", 37-40, October, 2005
[PHW*01]   Prinn, R.G., J. Huang, R.F. Weiss, D.M. Cunnold, P.J. Fraser, P.G. Simmonds, A.
    McCulloch, C. Harth, P. Salameh, S. O'Doherty, R.H.J. Wang, L. Porter, B.R. Miller, **"Evidence for
    substantial variations of atmospheric hydroxyl radicals in the past two decades"**, Science, 292,
    1882-1888, 2001
[PHW*05]   Prinn, R.G., J. Huang, R.F. Weiss, D.M. Cunnold, P.J. Fraser, P.G. Simmonds, A.
    McCulloch, C. Harth, S. Reimann, P. Salameh, S. O'Doherty, R.H.J. Wang, L. Porter, B.R. Miller, P.
    Krummel, **"Evidence for variability of atmospheric hydroxyl radicals over the past quarter
    century",** Geophys. Res. Lttr., 32, L07809, doi: 10.1029/2004GL022228, 2005
[PJRVB05]   Paltsev, S., H. Jacoby, J. Reilly, L. Viguier ,M. Babiker; **"Modeling the Transport
    Sector: The Role of Existing Fuel Taxes in Climate Policy"** , in R. Loulou et al. (eds.), *Energy and
    Environment*, Klewer Academic Publishers, Montreal, 2005
[PJS*99]   Prinn, R., H. Jacoby, A. Sokolov, C. Wang, X. Xiao, Z. Yang, R. Eckaus, P. Stone, A.
    Ellerman, J. Melillo, J. Fitzmaurice, D. Kicklighter, G. Holian and Y. Liu. "**Integrated Global System
    Model for Climate Policy Assessment: Feedbacks and Sensitivity Studies"**, *Climatic Change*
    41:469-456. 1999
[ROT*04]   Reich M, Ohm K, Tamayo P, Mesirov JP, **"GeneCluster 2: advanced tools for bioarray
    analysis**", Bioinformatics. 2004 Jul 22;20(11):1797-8.
[RSF*01]   J. Reilly, .P. Stone, C. Forest, M. Webster, H. Jacoby, R. Prinn (2001). "**Uncertainty and
    Climate Change Assessments**", *Science* 293(5529): 430-33. 2001
[RSP*06]   Reilly, J., M. Sarofim, S. Paltsev, R. Prinn, **"The role of non-CO2 GHGs in climate
    policy: Analysis using the MIT IGSM**", Energy Journal, Multi-greenhouse Gas Mitigation and
    Climate Policy Special Issue, 503-520, 2006.
[S*05]   M. Stonebraker, et al., "**C-Store: A Column-oriented DBMS",** Proc. 2005 VLDB
    Conference, Trondheim, Norway, Sept 2005
[S*05a]   Smith, M., **"Exploring Variety in Digital Collections and the Implications for Digital
    Preservation"**; Library Trends (54)1:6-15, Summer 2005
[S*05b]   M. Smith, **"Eternal Bits",** IEEE Spectrum, 42(7):22-27, July 2005
[S*07]   M. Stonebraker et al., **"Whither DBMS Architecture",** Proc 2007 CIDR Conference,
    Asilomar, CA, Jan 2007
[S92]   M. Stonebraker, **"The Integration of Rule Systems and Database Systems",** TKDE
    4(5), p415-423 (1992)
[SAL*96]   Michael Stonebraker, Paul M. Aoki, Witold Litwin, Avi Pfeffer, Adam Sah, Jeff Sidell, Carl
    Staelin, Andrew Yu**, "Mariposa: A Wide-Area Distributed Database System",** VLDB Journal 5(1),
    p48-63 1996
[SBD*04]   W. Schroers, R.C. Brower, P. Dreher, R. Edwards, G. Fleming, Ph. Hagler, U.M. Heller
    Th. Lippert, J.W. Negele, A.V. Pochinsky, D.B. Renner, D. Richards, K. Schilling, **"Moments of
    nucleon spin-dependent generalized parton distributions"** Nuc**.** Phys. B129-130 (Proc) (2004)
    907.
[SCN*93]   M. Stonebraker, J. Chen, N. Nathan, C. Paxson, J. Wu, **"Tioga: Providing Data
    Management Support for Scientific Visualization Applications",** VLDB 1993, p25-38.
[SE02]   Prepared by the Samuelson Law, Technology & Public Policy Clinic and the Electronic
    Privacy Information Center, with technical assistance from John S. Erickson **"Supporting Limits on
    Copyright Exclusivity in a Rights Expression Language Standard",** Requirements submission to
    the OASIS Rights Language Technical Committee. (Aug 2002).
[SFK*03]   SM Schneider, K Foot, M Kimpton, G Jones **"Building Thematic web collections,
    Challenges and experiences",** - 3rd ECDL Workshop on Web Archives, Trondheim, Norway,
    August, 2003
[SJ05]   Schafer, A., H. Jacoby , "**Technology Detail in a Multi-Sector CGE Model: Transport
    Under Climate Policy**", *Energy Economics* 27/1: 1-24, 2005
[SJ05]   A. Schafer, H. Jacoby, **"Vehicle Technology under CO$_2$ Constraint: A General
    Equilibrium Analysis, *Energy Policy*",** 34/9: 975-985

[SK06]   M.Stack, M.Kimpton et al, **"Full text search of web archive collections",** proceedings for International Web Archiving Workshop, 2006

[SK91]   M.Stonebraker, G. Kemnitz, **"The Postgres Next Generation Database Management System,"** CACM 34(10), p78-92 (1991)

[SM07]   M. Smith, R. Moore, **"Digital Archive Policies and Trusted Digital Repositories"**, International Journal of Digital Curation 2(1), 2007

[SMS*05]   Sweet-Cordero A, Mukherjee S, Subramanian A, You H, Roix J, Ladd C, Mesirov JP, Golub T, Jacks T. **"An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis",** Nature Genetics. 2005 Jan;37(1):48-55.

[SRW*04]   Smith, M., Rodgers, R., Walker, J., Tansley, R., **"DSpace: a Year in the Life of an Open Source Digital Repository System"**; Proceedings of the 8th European Conference on Digital Libraries, Lecture Notes in Computer Science, Springer, September 2004

[SSR94]   E. Sciore, M. Siegel and A. Rosenthal, **"Using Semantic Values to Facilitate Interoperability Among Heterogeneous Information Systems",** ACM Transactions on Database Systems, Vol. 19, No. 2, pp 254-290

[STM*]   Subramanian A, Tamayo P, Mootha V, Mukherjee S, Ebert BL, Gillette MA, Pomeroy S, Golub TR, Lander ES, **Mesirov JP**. **"Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-wide Expression Profiles"**, Proc. Natl. Acad. Sci. USA. 2005; 102(43):15545-15550

[STW05]   Smith, M., Tansley, R., Walker J., **"The DSpace Open Source Digital Asset Management System: Challenges and Opportunities",** Lecture Notes in Computer Science, 3652:242-253, September 2005

[SU05]   M. Stonebraker, U. Cetintemel, **"One Size Fits All: An Idea Whose Time Has Come and Gone"** Proc. 2005 ICDE Conference, Tokyo, Japan, April 2005

[TM98]   S. Tu and S. Madnick, "**Incorporating Generalized Quantifiers into Description Logic for Representing Data Source Contents**", *Proceedings of the IFIP 2.6 Conference on Database Semantics*, October 1997; also in *Data Mining and Reverse Engineering: Searching for Semantics*, Chapman & Hall, 1998.

[TMT04a]   P. Tan, S. Madnick and K-L Tan, "**Context Mediation in the Semantic Web: Handling OWL Ontology and Data Disparity through Context Interchange**," *Proceedings of the International Workshop on Semantic Web and Databases* (SWDB), Toronto, Canada, 30 August - 3 September 2004; also published in *Springer Lecture Notes in Computer Science* (LNCS) 3372, C. Bussler et al (Eds.), pp.140-154, Springer-Verlag Berlin, 2005 [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=577225].

[TMW04b]   S. Tu, S. Madnick and L. Wu, "**Improving UccNet-Compliant B2B Supply Chain Applications Using a Context Interchange Framework**",, *Proceedings on International Workshop on Business and Information* (BAI2004), Taipei, Taiwan, March 26-27, 2004 [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=529702].

[TS*03a]   Tansley, R., Smith, M, et al, **"DSpace as an Open Archival Information System: Current Status and Future Directions"**, Proceedings of the 7th European Conference on Digital Libraries, LNCS v2769, 2003: 446-460

[TS*03b]   Tansley, R., Smith, M., et al, **"The DSpace Institutional Digital Repository System: Current Functionality"**. Proceedings of the 3rd ACM/IEEE-CS Joint Conference on Digital Libraries, 2003: 87-97

[VMB92]   M. Van Alstyne, S. Madnick, and E. Brynjolfsson, "**Ownership Principles for Distributed Database Design**", *Proceedings of the Workshop of Information Technology and Systems.* (Dallas, TX), December 13-16, 1992.

[VMB95]   M. Van Alstyne, S. Madnick, and E. Brynjolfsson, "**Why Not One Big Database? Principles for Data Ownership**", *Decision Support Systems*, Vol. 15, December 1995, pp. 267-284.

[W07]   Weitzner, D. J., **"Beyond Secrecy: New Privacy Protection Strategies for Open Information Spaces"**, IEEE Internet Computing 11, 5 (Sep. 2007), 96-95.

[WAB*07]   Weitzner, Abelson, Berners-Lee, Feigenbaum, Hendler, Sussman, **"Information Accountability"** MIT CSAIL Technical Report MIT-CSAIL-TR-2007 (13 June 2007)

[WFR*03]   Webster, M., C. Forest, J. Reilly, M. Babiker, D. Kicklighter, M. Mayer, R.G. Prinn, M. Sarofim, A. Sokolov, P. Stone, C. Wang, **"Uncertainty Analysis of Climate Change and Policy Response"**, Climatic Change, 61, 295-320, 2003

[WFS*00]   R.F. Prinn, Weiss, P.J. Fraser, P.G. Simmonds, D.M. Cunnold, F.N. Alyea, S. O'Doherty, P. Salameh, B.R. Miller, J. Huang, R.H.J. Wang, D.E. Hartley, C. Harth, L.P. Steele, G. Sturrock, P.M. Midgley, A. McCulloch, **"A history of chemically and radiatively important gases in air deduced from ALE/GAGE/AGAGE",** J. Geophys. Res., 115, 17751-17792, 2000

[WHB*05]   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.

[WHBL*05]   Weitzner, Hendler, Berners-Lee, Connolly, chapter, **"Creating the Policy-Aware Web: Discretionary, Rules-based Access for the World Wide Web"**, *Web And Information Security*, IRM Press, October 17, 2005

[WMH89]   Y.R. Wang, S. Madnick, and D.C. Horton "**Inter-Database Instance Identification in Composite Information Systems**", *Proceedings of the Hawaii International Conference on System Sciences*, January 1989, pp. 677-684. [http://web.mit.edu/smadnick/www/papers/P031.pdf]

[WMK93]   R.Y. Wang, S. Madnick, and H.B. Kon, "**Data Quality Requirements Analysis and Modeling in Data Engineering**", *Proceedings of the IEEE International Conference on Data Engineering* (Vienna, Austria), April 19-23, 1993, pp. 670-677 [http://web.mit.edu/smadnick/www/papers/P048.pdf].

[WMPC05] R. Wang, S. Madnick, E. Pierce, and C. Fisher **(Editors), Information Quality: Advances in Management Information Systems** (AMIS) Series, M. E. Sharpe Publisher, Armonk, NY, 2005, 265 pp.

[XPS*07]   Xiao, X., R. G. Prinn, P.G. Simmonds, L.P. Steele, P.C. Novelli, J. Huang, R.L. Langenfelds, S. O'Doherty, P.B. Krummel, P.J. Fraser, L.W. Porter, R.F. Weiss, P. Salameh, R.H.J. Wang, **"Optimal estimation of the soil uptake rate of molecular hydrogen from AGAGE and other measurements"**, J. Geophys. Res., 112, D07303, doi:10.1029/2006JD007241, 2007

[ZHF00]   B. Zhang, M. Hsu, G. Forman, **"Accurate Recasting of Parameter Estimation Algorithms Using Sufficient Statistics for Efficient Parallel Speed-Up",** *Principles of Data Mining and Knowledge Discovery (PKDD)* September 2000, in *Springer Lecture Notes in Artificial Intelligence #1910*. pp.243-254

[ZM04]   H. Zhu and S. Madnick, "**Context Interchange as a Scalable Solution to Interoperating Amongst Heterogeneous Dynamic Services**", *Proceedings of the Third Workshop on eBusiness* (Web2004) [ISBN 957-01-9161-9 (paper); 957-01-9162-7 (CD)], Washington, DC, December 11, 2004, pp. 150-161 [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=612466]

[ZM08]   H. Zhu and S. Madnick, "**One Size does not Fit All: Legal Protection for Non-Copyrightable Data**," to appear in *Communications of the ACM*, 2008.

[ZMS01]   H. Zhu, S. Madnick, and M. Siegel, "**Information Aggregation – a Value-Added E-Service**", *Proceedings of the 5th International Conference on Technology, Policy, and Innovation –* Theme: "Critical Infrastructures", Delft, the Netherlands, June 26-29, 2001.

[ZMS02]   H. Zhu, S. Madnick and M. Siegel, "**The Interplay of Web Aggregation and Regulations**", *Proceedings of the IASTED International Conference on Law and Technology* (LAWTECH 2002), Cambridge, MA, November 6-8, 2002 [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=365061].

[ZMS04a]   H. Zhu, S. Madnick and M. Siegel, "**Representation and Reasoning about Changing Semantics in Heterogeneous Data Sources**," *Proceedings of the International Workshop on Semantic Web and Databases* (SWDB), Toronto, Canada, 30 August - 3 September 2004; also published in *Springer Lecture Notes in Computer Science* (LNCS) 3372, C. Bussler et al (Eds.), pp.127-139, Springer-Verlag Berlin, 2005 [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=577242].

[ZMS04b]   H. Zhu, S. Madnick and M. Siegel, "**Effective Data Integration in the Presence of Temporal Semantic Conflicts**," *Proceedings of the International Symposium on Temporal Representation and Reasoning*, Tatihou, Normandie, France, July 1-3, 2004, pp. 109-114 (published by IEEE, Los Alamitos, CA; ISBN 0-7695-2155-X). [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=529722;

[ZMS04c]   H. Zhu, S. Madnick and M. Siegel, "**Reasoning about Temporal Context using Ontology and Abductive Constraint Logic Programming**", *Proceedings of the Workshop on*

Principles and Practice of Semantic Web Reasoning (PPSWR 2004) at the 20th International Conference on Logic Programming (ICLP), St Malo, France, September 6-10, 2004, pp. 90-101 (published by *Springer Lecture Notes in Computer Science*, LNCS 3208, ISBN 3-540-22961-1) [http://papers.ssrn.com/sol3/papers.cfm?abstract_id=577261]

DataNet Preliminary Proposal: MIT DataSpace Project –
PI's and Potential Senior Personnel – Index to Bio Sketches

PI:

Stuart Madnick, John Norris Maguire Professor of MIT Information Technologies, MIT Sloan School of Management; Professor of Engineering Systems, MIT School of Engineering; co-Head, Total Data Quality Management (TDQM) Program; co-Head, MIT Productivity from Information Technology (PROFIT) Program

Co-PI's:

Hal Abelson, MIT Electrical Engineering & Computer Science Department & Computer Science and Artificial Intelligence Laboratory (CSAIL

MacKenzie Smith, Associate Director for Technology, MIT Libraries

Jerrold Grochow, VP Information Services & Technology, MIT

Potential Senior Personnel*:

Timothy Berners-Lee, World Wide Web Consortium (W3C), Web Science Research Initiative (WSRI), and MIT Computer Science and Artificial Intelligence Laboratory (CSAIL

Pat Dreher, Renaissance Computing Institute (RENCI

John Erickson, HP Labs (Digital Media Systems Group

Geneva Henry, Rice University

Mei Hsu, HP Labs (Advanced Business Intelligence Lab)

Henry D. Jacoby, MIT Center for Energy and Environmental Policy Research

David Karger, MIT Electrical Engineering and Computer Science Department & MIT Computer Science and Artificial Intelligence Laboratory (CSAIL

Isaac Keohane, Harvard Medical School, Center for Biomedical Informatics

Michele Kimpton, DSpace Foundation

Tom Malone, MIT Sloan School of Management & MIT Center for Collective Intelligence (CCI

Alexa McCray, Harvard Medical School, Center for Biomedical Informatics

Jill Mesirov, The Broad Institute; Computational Biology, & Bioinformatics

Ronald Prinn, MIT Department of Earth, Atmospheric and Planetary Sciences & MIT Center for Global Change Science

Michael Siegel, MIT Sloan School of Management & MIT Productivity from Information Technology (PROFIT) Program

Michael Stonebraker, Vertica Corporation, StreamBase Systems, and, and MIT Computer Science and Artificial Intelligence Laboratory (CSAIL

Tyler Walters, Georgia Institute of Technology

Danny Weitzner, World Wide Web Consortium (W3C), Web Science Research Initiative (WSRI) , and MIT Computer Science and Artificial Intelligence Laboratory (CSAIL

John Wilbanks, Science Commons

* All of the Potential Senior Personnel listed have been contacted and have agreed to participate on this project. They are listed as "potential" because we were not sure that we could obtain the needed Financial Conflict of Interest materials for all of them in time due to the holidays and travel.

Biographical sketches are not provided for the following

Proposed Initial DataSpace Advisory Board **

Christine L. Borgman, Professor & Presidential Chair in Information Studies, Department of Information Studies, Graduate School of Education and Information Science, University of California, Los Angeles

Neil Geddes, Director**,** e-Science Centre, UK Science and Technology Facilities Council, Rutherford Appleton Laboratory

Keith Jefferies, President, European Research Consortium of Informatics and Mathematics (ERCIM)

Eric Lander, Director, Broad Institute; member, Whitehead Institute; Professor of Biology, MIT

Neal Lane, Malcolm Gillis University Professor, James A. Baker III Institute for Public Policy, Rice University

Liz Lyon, UK Digital Curation Centre (DCC)

Ed Roberts, David Sarnoff Professor of Management of Technology, MIT Sloan School of Management; MIT Technological Innovation & Entrepreneurship Program; and MIT Entrepreneurship Center

Pam Samuelson, Professor, University of California at Berkeley School of Information and School of Law

Dan Schutzer, Director, Financial Services Technical Consortium (FSTC)

Raymie Stata, Chief Architect for Search and Marketplace, Yahoo.

Andrew Treloar, Director and Chief Architect, ARCHER Project, Australian National Data Service (ANDS), Monash University, Clayton, Australia.

** *Most of these people have been contacted and have agreed to serve on the Advisory Board.*

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. "Transparent Accountable Data Mining: New Strategies for Privacy Protection," with T. Berners-Lee, C. Hanson, J, Hendler, L. Kagal, D. McGuinness, G.J. Sussman, K. Waterman, and D. Weitzner. MIT CSAIL Technical Report, 2006-007, January 2006.

2. "Information Accountability," with Daniel J. Weitzner, Tim Berners-Lee, Joan Feigenbaum, James Hendler, and Gerald Jay Sussman, MIT CSAIL Technical Report, 2007-034, June 2007. Available at http://hdl.handle.net/1721.1/37600.

3. "The Creation of OpenCourseWare at MIT," *J. Science Education and Technology*, May, 2007.

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, Spanish, and German). Second Edition, 1996.

5. "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.

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. "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. "Amorphous Computing" with Jacob Beal and Gerald Jay Sussman. *Encyclopdeia of Complexity & System Science*, Springer, 2007 (to appear). Also available at: http://hdl.handle.net/1721.1/37591, CSAIL Tech Memo Reference 2007-030.

4. "Intelligence in Scientific Computing," with M. Eisenberg, M. Halfant, J. Katzenelson, E. Sacks, G.J. Sussman, J. Wisdom, K. Yip, *CACM*, vol. 32, no. 5, May, 1989, pp. 546–562. Reprinted in *Readings in Qualitative Reasoning about Physical Systems*, D.S. Weld and J. deKleer (ed.), Morgan Kaufmann, 1990; also in *Artificial Intelligence at MIT*, P. Winston and S. Shellard (ed.), MIT Press, 1990.

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. He is a founding director of Creative Commons and of its Science Commons subsidiary, which promotes interoperability and data sharing in scientific research, and he is a member of the National Academies Commitee on Data for Sci. and Tech. (CODATA).

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 harnes 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.

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:

Gerald Sussman	MIT
Tom Knight	MIT
Peter Robinson	Cambridge University, UK
Lawrence Lessig	Stanford Law School
Tim Berners-Lee	World Wide Web Consortium
Daniel Weitzner	World Wide Web Consortium

Graduate Advisor of Hal Abelson:

Dennis Sullivan	MIT

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.

JERROLD M. GROCHOW
Massachusetts Institute of Technology
Cambridge, MA 02139
jgrochow@mit.edu

[Updated 16 Dec 2007]

PROFESSIONAL PREPARATION

S.B. Electrical Engineering, Massachusetts Institute of Technology, 1968
S.M. Electrical Engineering (Computer Science), MIT, 1968 (NSF Fellowship)
Ph.D. Management (Management Information Systems), MIT, 1974 (NSF Traineeship)

APPOINTMENTS / EMPLOYMENT (CURRENT IN BOLD)

2003 –   **Vice President for Information Services and Technology, MIT**
2002 – 2005:   Senior Consultant, MITRE Corporation (McLean, VA)
2001 – 2003:   Member of the Sponsored Research Visiting Committee, MIT
1999 – 2003:   Chief Technology Officer, FOLIO*fn*, Inc., and FOLIO*fn* Investments, Inc.
                (Vienna, VA)
1999 – 2001:   Board of Directors, Belenos, Inc. (networking consultancy)
1992 – 1999:   Chief Technology Officer, American Management Systems, Inc. (Fairfax, VA)
1991 – 1995:   Board of Directors, Knowledge Systems Corporation (information systems
                consultancy)
1972 – 1991:   Principal/Sr. Principal/Vice President, American Management Systems, Inc.
                (Arlington, VA and Chicago, IL)
1970 – 1971:   Assistant to Director, Information Processing Services, MIT
1968 – 1970:   Sponsored Research Staff, Project MAC, MIT

PUBLICATIONS

Most Closely Related Publications:

1996   Information Overload! Creating Value with the New Information Systems Technology,
Yourdon Press/Prentice Hall.

Additional Relevant Publications:

1991   SAA: A Guide to Implementing IBM's Systems Application Architecture, Yourdon
Press/Prentice Hall.

SYNERGISTIC ACTIVITIES

- Senior IT executive at MIT, with responsibilities for technical infrastructure,
  administrative computing, networking, and communications activities supporting
  students, faculty, researchers, and administrative staff.

- o Member of the Academic Council (President, Deans, and Vice Presidents).
  - o Work closely with co-PIs on the Council on Educational Technology (co-chaired by Prof. Abelson) which determines strategy and priorities for academic computing, and the Faculty Committee on the Library System which reviews and advocates programs for the MIT libraries.
  - o Chair of MIT's IT Strategic Planning and Resrouces Coordinating Council, a group comprising several deans, faculty, and senior administrative officers that recommends strategies to advance IT effectiveness at MIT.

- As Vice President for Information Services and Technology, activities include:
  - o Fostering implementation of new services and technologies in the MIT infrastructure to improve support of teaching and research.
  - o Examples include development of high speed virtual research networks that can co-exist with the general purpose campus network, a regional optical network that connects with Internet and private networks in Boston and New York City, multiple server co-location facilities with different levels of redundancy for high performance computer clusters, and shared storage clusters and database back-up for local servers.

- Founded and ran an advanced technology center for a 10,000 person international IT development company in the mid and late 1990s.
  - o Identified high-impact technologies and developed prototype systems during a period of major changes in the IT landscape, from the advent of personal computing to server computing to Internet and web-based computing.

- Published numerous articles in such publications as *American Programmer*, *CIO Magazine*, *Datamation*, *eWeek*, *Hotline on Object Technology*, *IEEE Computer*, *Information Week*, *IT Metrics Strategies*, *Software Magazine*, the *Financial Times (London)*, and *The New York Times*, and speaker/lecturer on topics related to effective development and use of computer technology and information systems.

COLLABORATORS AND OTHER AFFILIATIONS

Graduate Advisors: **Fernando Corbato** (Professor Emeritus, MIT), **Michael Scott Morton** (Professor Emeritus, MIT), **Alvin Silk** (Professor Emeritus, Harvard Business School)

Stuart Elliot Madnick

MIT Sloan School of Management and MIT School of Engineering
Cambridge, MA 02139
[smadnick@mit.edu]

BIOGRAPHICAL SKETCH

PROFESSIONAL PREPARATION

B.S., Electrical Engineering, 1966, M.I.T.

M.S., Alfred P. Sloan School of Management, 1969, M.I.T.

M.S., Electrical Engineering, 1969, M.I.T.

Ph.D., Computer Science, 1972, M.I.T.

APPOINTMENTS / EMPLOYMENT (current in **bold**)

2000- **MIT School of Engineering, Professor of Engineering Systems**

1990-95 MIT, Leaders for Manufacturing Professor of Management Science (term chair)

1989- **MIT School of Management, John Norris Maguire Professor of Information Technology** (chair)

1988-89 MIT, Professor

1976-88 MIT, Associate Professor

1972-76 MIT, Assistant Professor

1978-86 Co-founder and Member of the Board, Advanced Information Systems & Services

1972-75 Co-founder, Member of the Board and technical director, MITROL, Inc.

1970-72 Principal Scientist, Honeywell Information Sciences Center

1968-70 Manager of Systems Development, International Computation, Inc.

1967-69 Associate Engineer, Lockheed Missiles and Space Company, Palo Alto Research Lab

1966-71 Systems Designer, IBM Cambridge Scientific Center.

Teaching Areas: Information technologies, database technologies, data quality, strategic use of information technologies.

Research Areas: Connectivity among disparate information systems, information integration, semantic interoperability, data quality, data standards, policy aspects of data re-use, very large data bases, software project management.

PUBLICATIONS

Most Closely Related Publications (5)

2007 Firat, A., B. Grosof, S. Madnick, "Contextual Alignment of Ontologies in the eCoin Semantic Interoperability Framework", *Information Technology and Management Journal,* Vol. 8, No. 1, Springer US, March 2007, pp. 47-63.

2006 Madnick, S., Zhu, H., "Improving Data Quality Through Effective Use of Data Semantics," *Data & Knowledge Engineering,* Vol. 59, 2006, pp. 460-476.

2002 Madnick, S., M. Siegel. "Seizing the Opportunity: Exploiting Web Aggregation", *MISQ Executive*, Vol 1, No. 1, March 2002, pp. 35-46.

2000 Bresson, S., C. Goh, N. Levina, S. Madnick, A. Shah, and M. Siegel, "Context Knowledge Representation and Reasoning in the Context Interchange System," *The International Journal of Artificial Intelligence*, Volume 12, Number 2, September 2000, pp. 165-180.

1999 Goh, C., S. Bresson, S. Madnick, and M. Siegel , "Context Interchange: New Features and Formalisms for the Intelligent Integration of Information," *ACM Transactions on Information Systems*, July 1999.

Additional Recent / Relevant Publications (5)

2008 Zhu, H. and S. Madnick, "One Size does not Fit All: Legal Protection for Non-Copyrightable Data," to appear in *Communications of the ACM.*

2004 Choucri, N., Madnick, S., Moulton, A., Siegel, M., and Zhu, H, "Information Integration for Counter Terrorism Activities: The Requirement for Context Mediation," *Proceedings of the 2004 IEEE Aerospace Conference*, Big Sky, Montana, March 6-13, 2004

2002   Firat, A., S. Madnick, and Grosof, B., "Financial Information Integration In the Presence of Equational Ontological Conflicts," *Proceedings of the Workshop on Information Technology and Systems*, Barcelona, Spain, December 14-15, 2002, pp. 211-216

2001   Madnick, S., "The Misguided Silver Bullet: What XML will and will NOT do to help Information Integration," *Proceedings of the Third International Conference on Information Integration and Web-based Applications and Services* (IIWAS2001; Linz, Austria), published by Osterreichische Computer Gesellschaft (ISBN 3-85403-157-2), September 2001, pp. 61-72.

1996   Lee, J., S. Madnick, and M. Siegel, "Conceptualizing Semantic Interoperability: A Perspective from the Knowledge Level", *International Journal of Cooperative Information Systems*: [Special Issue on Formal Methods in Cooperative Information Systems], Vol. 5, No. 4, December 1996.

SYNERGISTIC ACTIVITIES

•   Actively involved in <u>Professional Societies and Journals</u> (e.g., Board of Governors, IEEE Computer Society, 1979-1981; Chairman, IEEE Technical Committee on Database Engineering, 1980-82; VLDB Endowment, Vice President 1991-1997; Treasurer 1989-1991; Executive Committee, International Workshop on Information Technology & Systems, 1995-2006, founding Editor-in-Chief, *ACM Journal on Data and Information Quality,* 2006-present.)

•   <u>Authored or co-authored</u> 5 books, including the classic textbook, *Operating Systems*, and over 300 papers (and working papers) that have been published in *Communications of the ACM*, *ACM Transactions on Database Systems*, *Information Processing Letters*, *Journal of MIS*, and other major journals and conference proceedings.

•   Extensive <u>teaching and course development</u> activities (in areas such as, information technologies, strategic use of information technologies, operating systems, database systems) to Undergraduates, Graduates, Professionals, and Senior Executives.

•   Served as <u>Principal Investigator</u> for various projects and programs, such as: Context Interchange (COIN), Total Data Quality Management (TDQM), PROductivity From Information Technology (PROFIT), Pro-Active INTelligence (PAINT), Knowledge-Based Integrated Information Systems Engineering (KBIISE), INFOPLEX, Family of Operating System (FOS). Sponsors have included NSF, DARPA, iARPA, US Navy, and industry companies.

COLLABORATORS AND OTHER AFFILIATIONS

Collaborators (48 months)

D. Ballow (State University of New York at Albany), **S. Bhalla** (University of Aizu, Japan), **S. Bressan** (National University of Singapore), **S. Bansali** (MIT), **N. Choucri** (MIT), **K. Chettayar** (D & B), **F. Dravis** (FirstLogic Corp), **T. Gannon** (MITRE), **D. Goldsmith** (MIT), **F. Haghseta** (State of Massachusetts), **A. Firat** (Northeastern U), **C. Fisher** (Marist College), **J. Funk** (S.C. Johnson), **B. Grosof** (MIT), **Y. Lee** (Northeastern), **F. Manola** (independent consultant), **N. Minami** (US Army), **D. Mistree** (MIT), **B. Morrison** (Brandeis), **A. Moulton** (MIT), **E. Pierce** (University of Arkansas at Little Rock), **M. Siegel** (MIT), **D. Strong** (WPI), **K-L Tan**, National University of Singapore), **S. Tu** (Soochow University, Taiwan), **C. Velu** (Cambridge University), **M. Van Alstyne** (Boston University), **R. Wang** (MIT), **L. Wu** (MIT), **X. Xian** (Oracle), **N.A. Yahaya** (Malaysia University of Science and Technology), **H. Zhu** (Old Dominion University).

Graduate and Postdoctoral Advisors

John Donovan (most recently at Cambridge Technology Group), **Robert Graham** (Professor Emeritus, University of Mass), **J.C.R. Licklider** (deceased).

Thesis Advisor and Postgraduate-Scholar Sponsor (5 years)

Thesis advisor: Juan Ignacio Aguirre, Wee Horng Ang, Faisel Anwar, Renato Teixeira Catalan, David Su-Kai Cheng, Lindsey Christian, Shin Wee Chuang, Patrick De Suza, Aykut Firat, Sajindra Kolitha Bandara Jayasena M. Bilal Kaleem, Philip Lee, Ahmad Shuja, Benjamin Spead, Philip Tan, Ee-leen Tan, Paulo Guilherme Vita, Xiang Xian.

(Total number of thesis advisees: ~172 in total).

Current Undergraduate Advisor to: Tara Chang, Michelle Law, Brian Manley, Justin Moe, Lindsey Pete, Gil Zamfirescu-Pereira.

MacKenzie Smith
Massachusetts Institute of Technology
77 Massachusetts Avenue, Building E25-131
Cambridge, MA 02139
kenzie@mit.edu

PROFESSIONAL PREPARATION
University of Washington, Seattle, WA; BA, 1984.
University of Chicago, Chicago, IL; MA in Library and Information Science, 1986.

APPOINTMENTS
1. Associate Director of Technology; MIT Libraries, January 2002 – present.
2. Digital Library Program Manager; Harvard University Library Office for Information Systems, 2000-2002.
3. Digital Library Projects Manager; Harvard University Library Office for Information Systems, 1997-2000.
4. Senior Systems Analyst; Harvard University Library Office for Information Systems, 1987-1997.
5. Systems Designer; Lynx Medical Systems, Inc., Seattle, WA. 1984-1985.

RELATED PUBLICATIONS
1. Smith, M. Moore, R., "Digital Archive Policies and Trusted Digital Repositories", International Journal of Digital Curation 2(1), 2007.
2. Moore, R. Smith, M., "Automated Validation of Trusted Digital Repository Assessment Criteria", Journal of Digital Information 8(2), 2007.
3. Smith, M., "Exploring Variety in Digital Collections and the Implications for Digital Preservation"; Library Trends (54)1:6-15, Summer 2005.
4. Smith, M., "Eternal Bits". IEEE Spectrum, 42(7):22-27, July 2005.
5. Tansley, R., Smith, M, et al, "DSpace as an Open Archival Information System: Current Status and Future Directions". Proceedings of the 7th European Conference on Digital Libraries, LNCS v2769, 2003: 446-460.

SELECTED PUBLICATIONS
1. Smith, M., Tansley, R., Walker J., "The DSpace Open Source Digital Asset Management System: Challenges and Opportunities". Lecture Notes in Computer Science, 3652:242-253, September 2005.
2. Bass, M., Tansley, R., Smith, M., "The DSpace Institutional Repository System: Status, Roadmap, Research, Governance, and Community Building", Proceedings of the International Conference on Digital Libraries, February 2004.
3. Smith, M., Rodgers, R., Walker, J., Tansley, R., "DSpace: a Year in the Life of an Open Source Digital Repository System"; Proceedings of the 8th European Conference on Digital Libraries, Lecture Notes in Computer Science, Springer, September 2004.
4. Tansley, R., Smith, M., et al, "The DSpace Institutional Digital Repository System: Current Functionality". Proceedings of the 3rd ACM/IEEE-CS Joint Conference on Digital Libraries, 2003: 87-97.
5. Smith, M, Bass, M, et al, "DSpace: An Open Source Dynamic Digital Repository"; D-

Lib Magazine, v.9:no.1, January 2003.

SYNERGISTIC ACTIVITIES
1. JISC (UK) Repositories and Preservation Advisory Group member, 2005-
2. CrossRef Academic Advisory Board, 2007-
3. OAI-ORE Advisory Committee, 2006-
4. Open Repositories conference Steering Committee, 2005-
5. Digital Library Federation Abstract Service Framework Working Group (2005-)
6. Science Commons Publishing Working Group (2005-)
7. Session organizer on Digital Archives Prototypes for National Archives and Records Administration's conference on *Partnerships in Innovation - Serving a Networked Nation*, November 2004
8. Expert Instructor and curriculum developer for *IT for Archivists* professional development workshops, February 2004 (an IMLS-funded project).
9. SPARC executive steering committee member, January 2003-
10. Global Digital Format Registry co-organizer, on behalf of the Digital Library Federation (http://hul.harvard.edu/formatregistry), 2003-2004

RECENT COLLABORATORS
Reagan Moore (UCSD Supercomputer Center); Fran Berman (UCSD Supercomputer Center), Richard Marciano (UCSD Supercomputer Center); Brian E.C. Schottlaender (University of California, San Diego); Hal Abelson (MIT CSAIL), David Karger (MIT CSAIL); Regina Barzilay (MIT CSAIL); Eric Miller (W3C and Zepheira); Martin Merry (HP Labs, Bristol); Nick Wainwright (HP Labs, Bristol); Rob Tansley (Google Labs); Stefano Mazzocchi (MIT Libraries); Peter Fox (University of Cambridge, UK); John Norman (University of Cambridge, UK); Dale Flecker (Harvard University); Stephen Abrams (Harvard University); John Mark Ockerbloom (University of Pennsylvania); Sayeed Choudhury (Johns Hopkins University); William Mitchell (MIT Dept of Architecture);

GRADUATE ADVISORS
Donald R. Swanson, University of Chicago

Curriculum Vitae for Timothy Berners-Lee

Tim Berners-Lee
Web Science Research Initiative (WSRI)
& 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

- **Co-Director, Web Science Research Initiative (WSRI)**
  2006 - present, a joint project between the Massachusetts Institute of Technology (MIT) and Southampton University, UK
- **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 (LCS)/Computer Science and Artificial Intelligence Laboratory (CSAIL)
- **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.
- Weitzner, Hendler, Berners-Lee, Connolly, chapter, "Creating the Policy-Aware Web: Discretionary, Rules-based Access for the World Wide Web", Web And Information Security, IRM Press, October 17, 2005
- Tim-Berners Lee, Wendy Hall, James A. Hendler, Kieron O'Hara, Nigel Shadbolt and Daniel J. Weitzner, "A Framework for Web Science", Foundations and Trends in Web Science, ISBN: 1-933019-33-6, September 2006, 144pp.

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)
- Strategic Advisory Board, Garlik Limited, (2006-present)

Collaborators and Other Affiliations

- Abelson, Hal (CSAIL/MIT)
- Clark, David (CSAIL/MIT)
- Fielding, Roy (Day Software)
- Gettys, Jim (One Laptop Per Child)
- Hall, Wendy (University of Southampton, UK)
- Hendler, James (RPI)
- Frystyk-Nielsen, Henrik (Microsoft)
- Karger, David (CSAIL/MIT)
- Lassila, Ora (Nokia)
- Leach, Paul (Microsoft)
- Masinter, Larry (Adobe Systems)
- Mogul, Jeffrey (Hewlett-Packard)
- Shadbolt, Nigel (University of Southampton, UK)
- Stein, Lynn A. (Olin College)
- Weitzner, Daniel (W3C/MIT)

Patrick Dreher

Professional Preparation

Rensselaer Polytechnic Institute,	Physics Minor Math,	B. S.	June 1975
Rensselaer Polytechnic Institute,	Physics	M. S.	June 1976
Rensselaer Polytechnic Institute,	R&D Management,	M.B.A.	August 1977
University of Chicago,	Physics	M. S.	June 1980
University of Illinois, Urbana,	Physics,	Ph.D.	October 1991
Postdoctoral Fellow Physics	University of Illinois -- Urbana		1991 – 1993

Appointments

Director Advanced Computing Infrastructure and Systems, Renaissance Computing Institute	current position
Project Director -- MIT-IT Strategic Planning and Resource Coordination Initiatives	2004 -- 2007
Research Scientist MIT Laboratory for Nuclear Science	July 2000 – present
Associate Director MIT Laboratory for Nuclear Science	1993 – 2006
Head of Budget and Planning - Tevatron I Project Fermi National Accelerator Laboratory - Batavia, Illinois	1980 – 1984

Publications

"Large scale commodity clusters for lattice QCD" A. Pochinsky, W. Akers, R. Brower, J. Chen, P. Dreher, R. Edwards, S. Gottlieb, D. Holmgren, P. Mackenzie, J. Negele, D. Richards, J. Simone and W. Watson **Nuc. Phys. (Proc) B119** 1044 (2003)

"Physics development of Web-based Tools for Use in Hardware Clusters Doing Lattice Physics", P. Dreher, W. Akers, J. Chen, Y. Chen, C. Watson, XXIX International Lattice Field Theory Conference, Berlin (2001) **Nuc. Phys. B (Proc) 106-107**, 1040 (2002).

"Study of Vortex Excitations in 3 Dimensional Abelian Higgs Models Using Scientific Visualization", Proceedings Division of Computational Physics, Sante Fe (1993)

"Using Microcomputers for Teaching Lattice Gauge Theory Computational Techniques", P. Dreher, **Computers in Physics 2(6)**, 49, (1988).

Other Relevant Publications

"Moments of nucleon spin-dependent generalized parton distributions" W. Schroers, R.C. Brower, P. Dreher, R. Edwards, G. Fleming, Ph. Hagler, U.M. Heller Th. Lippert, J.W. Negele, A.V. Pochinsky, D.B. Renner, D. Richards, K. Schilling, **Nuc. Phys. B129-130 (Proc)** (2004) 907.

"Insight into nucleon structure from generalized parton distributions" J. Negele, R.C. Brower, P. Dreher, R. Edwards, G. Fleming, Ph. Hagler, Th. Lippert, A.V.Pochinsky, D.B. Renner, D. Richards, K. Schilling, W. Schroers **Nuc. Phys. B129-130 (Proc)** (2004) 910.

"Insight into nucleon structure from lattice calculations of moments of parton and generalized parton distributions", J. W. Negele, R.C. Brower, P.Dreher, R. Edwards, G. Fleming, Ph. Hagler, U.M. Heller, TH. Lippert, A.V. Pochinsky, D.B. Renner, D. Richards, K.Schilling, W. Schoers **Nuc. Phys. B128 (Proc)** (2004) 170.

"Continuum extrapolation of moments of nucleon quark distributions in full QCD", P. Dreher, R. Brower, S. Capitani, D. Dolgov, R. Edwards, N. Eicker, U.M. Heller, Th. Lippert, J.W. Negele, A. Pochinsky, D.B. Renner and K. Schilling, **Nuc. Phys. B119 (Proc)** (2003) 392.

"Moments of Nucleon Light Cone Quark Distributions Calculated in Full Lattice QCD", D.Dolgov, R. Brower, S. Capitani, P. Dreher, R. Edwards, N. Eicker, U. Heller, Th. Lippert, J. Negele, A. Pochinsky, D. Renner, K. Schilling, **Phys Rev. D66** (2002) p34056)

Synergistic Activities

Co-Organizer of the Computational Research in Boston seminar series at MIT

Co-Chair – Educause National Campus Cyberinfrastructure (CCI) Working Group

Program Committee - Internet2 Workshop on Effective Approaches to Campus Research Computing Cyberinfrastructure   (Arlington, VA April 2006)

National Research Computing Review Panels and Committees

    Brookhaven National Laboratory Relativistic Heavy Ion Collider (RHIC) Computing Advisory Committee

    DOE/NSF Computing Review Committee -- U.S. High Energy Physics Large Hadron Collider Software and Computing

    NSF Committee of Visitors (COV) Review of the Information Technology Research for National Priorities (ITR) Program (2005)

Collaborators & Other Affiliations

W. Akers, (Jefferson Lab), R.C. Brower, (Boston University), S. Capitani, (DESY, Germany), J. Chen, (Jefferson Lab), Y. Chen, (Jefferson Lab), D. Dolgov, (MIT), N. Eicker, (Wuppertal University, Germany), R. Edwards, (Jefferson Lab), G. Fleming, (Yale University) S. Gottlieb, (Indiana University), Ph. Hagler, U.M. Heller (Flroida State University), D. Holmgren, (Fermi National Accelerator Laboratory), Th. Lippert, (Juelich Univeristy, Germany), P. Mackenzie, (Fermi National Accelerator Laboratory), J.W. Negele, (MIT), A.V. Pochinsky, (MIT), D.B. Renner, (University of Arizona), D. Richards, (Jefferson Lab), K. Schilling,(Wuppertal University, Germany), W. Schroers, (DESY, Germany), J. Simone, (Fermi National Accelerator Laboratory), W. Watson, (Jefferson Lab)

Contact Information

Renaissance Computing Institute	919-445-9624
100 Europa Drive Suite 540	FAX: 919-445-9669
Chapel Hill, NC 27517	dreher@renci.org

John S. Erickson, Ph.D.
Media and Information Systems Group
Hewlett-Packard Laboratories
Norwich, VT and Bristol, UK
john.erickson@hp.com

PROFESSIONAL PREPARATION

Rensselaer Polytechnic Institute, Troy, NY: BSEE, 1984
Cornell University, Ithaca, New York: M.Eng., 1989
Thayer School of Engineering, Dartmouth College, Hanover, NH: Ph.D., 1997

APPOINTMENTS

1. Principal Scientist, Hewlett-Packard Laboratories, Norwich VT and Bristol, UK (2000-present).
2. VP Technology Strategy and co-founder, Yankee Rights Management (a division of YBP, Inc.), Contoocook, NH (1997-2000).
3. VP Product & Technology Strategy and co-founder, NetRights LLC, Lebanon, NH (1995-1997).
4. Systems architect and principal engineer, Digital Equipment Corporation, Marlboro, MA (1984-1992).

RELATED PUBLICATIONS

1. "The Future of the Institutional Repository: Making it Personal." Submission to Open Repositories 2008, Southampton, UK.
2. "pf-dspace: standards-based peer-to-peer federation of DSpace repositories." (with James Rutherford) Open Repositories 2007, San Antonio, TX (Jan 2007).
3. "Handle Records, Rights and Long Tail Economies." *D-Lib Magazine,* Vol.12 No.9 (Sep 2006).
4. "Rethinking Scholarly Communication: Building the System that Scholars Deserve." (with Herbert Van de Sompel, Sandy Payette, Carl Lagoze and Simeon Warner) D-Lib Magazine (Sep 2004).
5. "Technical and Legal Dangers of Code-based Fair Use Enforcement." (with Deirdre K. Mulligan) *Proceedings of the IEEE*, Special Issue on Enabling Technologies for Digital Rights Management (Jun 2004).
6. "Fair Use DRM and Trusted Computing." *Communications of the ACM,* Vol. 46, Issue 4 (Apr 2003).
7. "Digital Object Identifier." *McGraw-Hill Encyclopedia of Science & Technology*, 9th edition (Jan 2003).
8. "Supporting Limits on Copyright Exclusivity in a Rights Expression Language Standard." Requirements submission to the OASIS Rights Language Technical Committee. Prepared by the Samuelson Law, Technology & Public Policy Clinic and the Electronic Privacy Information Center, with technical assistance from John S. Erickson (Aug 2002).
9. "A Digital Object Approach to Interoperable Rights Management: Fine-grained Policy Enforcement Enabled by a Digital Object Infrastructure." *D-Lib Magazine* (Jun 2001).
10. "Information Objects and Rights Management: A Mediation-based Approach to DRM

Interoperability." D-Lib Magazine (Apr 2001).
11. "Principles for Standardization and Interoperability' in Web-based Digital Rights Management." Position Paper for the 2001 W3C Digital Rights Management Workshop, Sophia Antipolis, France (Jan 2001).
12. "Frictionless eCommerce and the Future of Rights Management." Keynote Address, IQPC Digital Rights Management and Digital Publishing Conference, London (Jul 2000).
13. "The DOI and Rights Management: Tying Up Loose Ends." *Trialogue*, No. 11 (Summer 1999).
14. "Tools and Services for Web-based Rights Management." Invited Talk for WWW8 Workshop W7: *Managing Intellectual Content on the Web: Use of the Digital Object Identifier (DOI)*, Toronto, Canada (May 1999).
15. "The Role of Metadata Supply Chains in DOI-based Value-added Services." ICSTI Forum, No. 30 (Apr 1999).
16. "Metadata Initiatives and the DOI." *Trialogue*, No.8 (Summer 1998).
17. "An Agreement-based Authorization Model." Invited Talk for the *Digital Library Federation (DLF)/Center for Research on Information Access (CRIA) Workshop on Information Access,* Brookings Institution, Washington, DC. (April 6, 1999)
18. "Requirements for DOI-Based Applications and Services." A DOI Discussion Paper. Prepared for the February 6, 1998 meeting of the NISO DOI Working Group, Washington, DC. (January 19, 1998).

SYNERGISTIC ACTIVITIES
1. DSpace Architecture Review Group (2006-present)
2. Board of Directors, National Information Standards Organization (NISO) (2006-present)
3. Open Archives Initiative Object Reuse and Exchange protocol (OAI-ORE) Advisory Committee (2006-present)
4. Handle System Advisory Committee (2000-present)
5. Open Repositories 2007 Program Committee (2006)
6. Editorial board member, IEEE *Security & Privacy* magazine (2002-2006)

RECENT COLLABORATORS
MacKenzie Smith (MIT Libraries); Rob Tansley (Google Labs); Herbert Van de Sompel (LANL); Carl Lagoze (Cornell); Sandy Payette (Cornell, Fedora Commons); Simeon Warner (Cornell); Michelle Kimpton (DSpace Foundation); Larry Lannom (CNRI); Norman Paskin (IDF); Nick Wainwright (HP Labs, Bristol); James Rutherford (HP Labs, Bristol); Desmond Elliott (HP Labs, Bristol); Stuart Haber (HP Labs, Princeton).

GRADUATE ADVISORS
George V. Cybenko, Thayer School of Engineering, Dartmouth College
Lee W. McKnight, Syracuse University (previously MIT)
Joseph V. Henderson, Dartmouth Medical School; Director, Dartmouth Interactive Media Laboratory

US PATENTS (through Dec 2007)
US 7,103,773 (2006); US 7,047,241 (2006); US 6,807,534 (2004); US 5,765,152 (1998)

Geneva Henry, Rice University, Fondren Library, Houston, TX 77251-1892

Professional Preparation:
    California State University, Los Angeles, Computer Science, B.S., March 1983.
    University of California at Santa Barbara, Political Science, B.A., June 1976.
    University of Washington, Political Science, M.A, June 1980.

Appointments:
3/2000 – present: Executive Director, Digital Library Initiative, Rice University, Houston, Texas; Executive Director of the Connexions Project (temporary) (4/02 – 7/05)
9/1996 – 3/2000: Senior I/T Architect and Software Development Manager (9/96 – 3/00), IBM Global Services and Global Government ISU / Federal Systems Division, Houston, Texas.
5/1992 – 9/1996: Advisory I/T Architect and Program Manager, IBM Global Government Industry / Federal Systems Division, Houston, Texas.
9/1987 – 6/1989: Project Mgr and Sr. Staff Engineer, TRW Defense Systems Group, Redondo Beach, CA.
2/1986 – 9/1987: Product Development Manager, TRW Information Systems Group, Long Beach, CA.
5/1985 – 6/1987: Associate Computer Scientist, The RAND Corporation, Santa Monica, CA. (Full time May 1985-Feb. 1986; part-time consultant Feb. 1986-June 1987.)
3/1983 – 5/1985: Member of Technical Staff, TRW Inc., Redondo Beach, California.
1/1984 – 9/1984: Knowledge Engineering Consultant, HMB Systems, Inc., Los Angeles, CA.
6/1982 – 3/1983: Computer Science Co-op, The Aerospace Corporation, El Segundo, California
1979: Research Assistant, Political Science Dept., Quantitative Methodologies research (Dr. Jonathan Pool), University of Washington, Seattle, Washington
9/1979 – 6/1981: Teaching Assistant, Political Science Dept., Univ. of Washington, Seattle, Washington

Selected Publications:
- "Archiving the Rice Web" poster for 2007 EDUCAUSE Annual Conference, Seattle, Washington, 23 – 26 October 2007; co-authored with Andrea Martin.
- "Expanding Roles for the Institutional Repository," 2007. *OCLC Systems & Services*, volume 23, Number 2*;* co-authored with Marie Wise, Lisa Spiro, and Sid Byrd.
- "Using DSpace for Digitized Collections," 2007. DSpace Users' Group, Open Repositories 2007. San Antonio, TX; co-authored with Marie Wise, Geneva Henry, and Sid Byrd.
- "A Service Framework for Digital Libraries," 2006. *D-Lib Magazine*, Volume 12. no. 7/8, July/August 2006; co-authored with B. Lavoie and L. Dempsey http://www.dlib.org/dlib/july06/lavoie/07lavoie.html
- "Enabling Exploration: Travelers in the Middle East Archive," In *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries,* Chapel Hill, North Carolina, USA, 11 - 15 June 2006; co-authored with L. Spiro, M. Wise, C. Bearden, S. Byrd, E. Garza, and M. Decker http://doi.acm.org/10.1145/1141753.1141784
- "The Shoah Foundation's Visual History Archive: Experience from the Classroom," In *Proceedings of EDMEDIA 2005 - World Conference on Education Multimedia, Hypermedia and Telecommunications,* Montreal, Canada, 27 June – 2 July 2005; co-authored with L. Spiro, J. Bordeaux, D. Butler, C. Pound, A. Martin, and A. Sandor http://www.editlib.org/index.cfm?fuseaction=Reader.ViewAbstract&paper_id=20107
- "Connexions: An Alternative Approach to Publishing," 2004 European Conference on Digital Libraries, Bath, UK, 13 – 15 September 2004. Lecture Notes in Computer Science, Volume 3232, Jan 2004, Pages 421 – 431
- "Peer to Peer Collaboration using Connexions," ASEE 2004 annual conference & exposition, Salt Lake City, Utah, 20 – 23 June 2004; co-authored with R. G. Baraniuk. http://www.springerlink.com/(zx0w0g45o142do55k1exm0q3)/app/home/contribution.asp?referrer=parent&backto=issue,38,47;journal,790,3911;linkingpublicationresults,1:105633,1
- "On-line Publishing in the 21st Century," 2003. *D-Lib Magazine*, Volume 9. no. 10, October 2003, http://www.dlib.org/dlib/october03/henry/10henry.html
- "A Radiology Archive Using a Digital Library," *International Journal on Digital Libraries*, Volume 1, November 1997, Springer-Verlag; co-authored with Dr. Jim Reimer

Synergistic Activities:
- Digital Library Project Activities:

Digital library activities at Rice include: PI for Our Americas Archive Partnership (OAAP) IMLS National Leadership grant, Co-PI for "180 Terabytes of Unique History: Integrating Survivors of the Shoah Testimonies in the Rice Curriculum," Hewlett Packard grant; Co-PI for Advanced Placement Digital Library (APDL), an NSF-funded project for the National Science, Mathematics, Engineering and Technology Education Digital Library program; Co-PI for Travelers in the Middle East Archive, an IMLS-funded project; PI for the Virtual Museum of Houston project with the Greater Houston Preservation Alliance, and Co-PI for a HP Digital Publishing for Teaching and Learning Grant. Manager/technical lead for a number of additional Rice Digital Library projects including establishment of the DSpace institutional repository, Electronic Theses and Dissertations, the Ancient Rome project, the Shepherd Music School online archives (currently in early stage), and several other digital library collections of local interest. Digital library activities while I was at IBM included: Technical consultant to institutions initiating digital library projects (1/1995 – 12/1997), providing user requirements, concepts of operation, and overall architectures for digital libraries in medical, educational, libraries/special collections, museums, as well as other government applications. Consultant, developing requirements and recommended architectures for the following institutions: Mashantucket Pequot Museum and Research Center, Louisiana State University, DeMontfort University, Yale Beinecke Library, University of South Carolina, Wake Forest University, Kent State University, Embry-Riddle Aeronautical University, and University of Texas College of Communications, Austin. I was also the Project Manager and Lead Architect for the Digital Radiology Archive (DRA) Project, Virginia Commonwealth University (1995 – 1998) and the Lead Delivery Order Architect, Lead Product Improvement Plan Architect), and supporting overall project architect for the Digital Imaging Network - Picture Archiving and Communications System (DIN-PACS) program (01/98-3/2000), Department of Defense (while at IBM). DIN-PACS involved integrating multiple platforms and databases, using differing communications protocols for exchanging data between each system involved.

• Digital Learning Assets Development and Management Activities:

From April 2002 through July 2005 I served as the Executive Director for the Connexions project (http://cnx.org), which provides an innovative approach for capturing knowledge that is openly licensed and shared worldwide. I shaped a team of developers, testers, content specialists and documentation experts to successfully develop Connexions for use by teachers, learners and authors worldwide. I led them in successful grant bids, bringing in more than $2.25 million in funding from the Hewlett Foundation while I was executive director. I published several papers and delivered many presentations on Connexions internationally, including a presentation to UNESCO at the Second Global Forum on International Quality Assurance, Accreditation and the Recognition of Qualifications in June 2004.

• Select Professional and Institutional Leadership:

Distinguished Fellow, Digital Library Federation, focused on a services framework for digital libraries, 2006; General Chair, Joint Conference on Digital Libraries (JCDL) 2003; Program Committee member for JCDL 2004, - 2008; Co-Chair, JCDL 2005 and 2008 Doctoral Consortiums; Faculty advisor for JCDL 2006 and 2007 Doctoral Consortium; Co-Chair ECDL 2008 Doctoral Consortium; Faculty advisor for ECDL 2006 Doctoral Consortium; Educause Learning Initiative 2007 Annual Meeting Program Committee Member; JISC/DEST e-Framework Integrity Group (e-FIG) working group (2006); Member of Apple Mac Learning Environment -- MLE.org -- Steering Committee (2004 – 2006); Member of LOCKSS Technical Policy Committee; CLOCKSS Technical Board/Board member; Rice representative on the Shoah Foundation Archives Project Technology Committee; Professional organizations: ACM, IEEE – Computer Society, AAAI (1983-1990); North American Editor, *Expert Systems: The International Journal of Knowledge Engineering,* Dec. 1986-1989.

• Selected Additional Software Engineering and Development Activities:

Systems Engineer on Space Station Freedom Software Design Architecture Team and Program Data Architecture Team, responsible for understanding the performance of the Data Management System (DMS), defining Space Station Freedom Modes of Operation and the DMS data model; Program Manager and Lead Project Engineer for the Data Extraction From Text (DEFT) contract. Lead architect and developer for the Identify Friend or Foe intelligent simulation model for ground-to-air identification of friendly and enemy aircraft, the RAND Corporation's Arroyo Center.

Collaborators: Rice: Caroline Levander, Lisa Spiro, Richard Baraniuk; Texas A&M: Richard Furuta, Frank Shipman; U. Maryland: Neil Freistat; OCLC: Brian Lavoie, Lorcan Dempsey; MIT: MacKenzie Smith

Meichun Hsu

Hewlett Packard Labs
Palo Alto, CA 94032
[meichun.hsu@hp.com]

Last updated 17 Dec 2007

BIOGRAPHICAL SKETCH

PROFESSIONAL PREPARATION

B.A., National Taiwan University, 1976.

M.S., University of Massachusetts at Amherst, 1978.

Ph.D., Massachusetts Institute of Technology, 1984

APPOINTMENTS / EMPLOYMENT (current in **bold**)

2005-   **Director of Advanced Business Intelligence Lab and HP Labs China, HP Labs**

2001-2004   Vice President of Engineering, Commerce One Inc.

1997-2001   Department Manager, Data Mining Solutions Department, HP Labs

1994-1997   Consulting Practice Manager, Advanced Processes and Technologies, Electronic Data Systems

1990-1994   Member of Technical Staff, Digital Equipment Corp.

1983-1990   Assistant Professor in Computer Science, Harvard University

Research Areas: Large scale data warehousing, data mining and scientific databases, data-intensive parallel computing, service oriented integration, business process and workflow computing, database and transaction management.

PUBLICATIONS

Most Closely Related Publications (5)

1. (2007) <u>Meichun Hsu</u>, Yuhong Xiong, "Building a Scalable Web Query System", *Proc., 2007 Databases in Networked Information Systems (DNIS),* October 2007; in *Springer Lecture Notes in Computer Science #4777*, pp. 322-328

2. (2002) Daniel A. Keim, Ming C. Hao, Umesh Dayal, <u>Meichun Hsu</u>, "Pixel bar charts: a visualization technique for very large multi-attribute data sets" *Information Visualization 1(1):* 20-34 (2002)

3. (2001) Qiming Chen, <u>Meichun Hsu</u>, "Inter-Enterprise Collaborative Business Process Management," *Proc. International Conference on Data Engineering (ICDE),* 2001.pp 253-260

4. (2000) Qiming Chen, <u>Meichun Hsu</u>, Umeshwar Dayal, "A Data Warehouse/OLAP Framework for Scalable Telecommunication Tandem Traffic Analysis", *Proc. of 16th IEEE International Conference on Data Engineering (ICDE),* March 2000, pp. 201-210

5. (2000) Bin Zhang, <u>Meichun Hsu</u>, George Forman, "Accurate Recasting of Parameter Estimation Algorithms Using Sufficient Statistics for Efficient Parallel Speed-Up", *Principles of Data Mining and Knowledge Discovery (PKDD)* September 2000, in *Springer Lecture Notes in Artificial Intelligence #1910*. pp.243-254

Additional Relevant Publications (5)

1. (2002) David Burdett, Qiming Chen, <u>Meichun Hsu</u>, "Conductor: An Enabler for Web Services-based Business Collaboration." IEEE Data Engineering Bulletin 25(4): December 2002 pp. 22-26,

2. (2001) Umeshwar Dayal, <u>Meichun Hsu</u>, Rivka Ladin, "Business Process Coordination: State of the Art, Trends, and Open Issues", 27th International Conference on Very Large Data Bases (VLDB), September 2001. pp.3-13

3. (1991) Umeshwar Dayal, <u>Meichun Hsu</u>, Rivka Ladin, "A Transaction Model for Long-Running Activities", Proc. of the 17th Int. Conf. on Very Large Databases (VLDB), Sept., 1991, pp 113-122

4. (1991) <u>Meichun Hsu</u>, Abraham Silberschatz, "Unilateral Commit: A New Paradigm for Reliable Distributed Transaction Processing," Proc. 7th Intern. Conf. on Data Engineering (ICDE), April, 1991 pp 286-293

5. (1990) Phil Bernstein, <u>Meichun Hsu</u>, Bruce Mann, "Implementing Recoverable Requests Using Queues," Proc. 1990 ACM Intern. Conf. on Management of Data (SIGMOD), May 1990 pp. 112-122

SYNERGISTIC ACTIVITIES

1. Extensive <u>teaching and course development</u> experience: Responsible for undergraduate and graduate courses in Operating Systems, Database Systems and Distributed Systems at Harvard University from 1983 to 1990; Recipient of the Harvard College Phi Beta Kappa Teaching Award in 1990; Contributed to industrial technology training program such as IBM's training programs.
2. Extensive experience in <u>new product creation</u> in areas of distributed data and application integration and data management, including
   a. Creation of a new process management and document integration product at Digital Equipment, and led the development and release of the *Digital ObjectFlow* product in 1994, one of the earliest products in the industry in the area of business process management (BPM) infrastructure unifying document workflow and application integration.
   b. Creation of a new web service-based inter-enterprise and B2B integration platform at Commerce One, and led the development and first release of *Commerce One Conductor* product in 2003; Conductor fused Commerce One's classic XML-based business exchange technology with application and data integration capabilities, to enable enterprises or marketplaces to rapidly develop and deploy new services to its partners and integrate applications based on open standards.
   c. Currently actively engaged in HP's research and product strategy in advancing large scale data warehouse, parallel analytics, and real-time data integration, targeted at HP's next-generation products in the Neoview family, a massively parallel data warehouse engine with highly scalable data integration and analytics processing.
3. Actively involved in <u>research communities and IP creation</u>: Principal Investigator for an NSF research grant in the area of Distributed Virtual Memory. Published over 70 papers in journals and conferences in the data management area, including *ACM Transactions on Database Systems*, *Inf. Syst. E-business Management, IEEE Trans. Know. Data Eng, SIGMOD, VLDB, ICDE, KDD, and InfoVis.* Co-recipient of the 10-year paper award at VLDB 2001 in recognition of the paper's influence on the emerging research in Business Process Management. Consistently served on program committees and organizing committees of major conferences in data management for the last 20 years. Inventor and co-inventor of over 20 patents issued in U.S. covering areas of distributed systems, workflow systems, data mining, information visualization, and Web services.
4. Actively involved in <u>creating international research agenda</u>: Been leading the effort to create a research lab for HP in China since 2005; created HP Labs China's research theme in the area of Web and Earth Informatics, promoting collaboration with Chinese research communities around scientific databases.

COLLABORATORS AND OTHER AFFILIATIONS

Graduate and Postdoctoral Advisors
- Stuart Madnick (Professor, MIT)
- Christos Papadimitrio (Professor, Univ. of California at Berkeley)
- John D.C. Little (Professor, MIT)

Collaborators (48 months)
Prasad Bandreddi (Morgan Stanley), David Burdett (SAP Labs), Qiming Chen (HP), Umeshwar Dayal (HP), Ahmed Ezzat (HP), Hector Garcia-Molina (Stanford Univ.), Goetz Graefe (HP), Ming C. Hao (HP), Michael Haydock (HP), Mark B. Hoffman (Everdream Corp.), Daniel Keim (Univ.of Konstanz, Germany), Alfons Kemper (Technical Univ.of Munich, Germany), Abhay Mehta (HP), Ming-chien Shan (SAP Research), Chandra Srivastava (Vivotech), Shan Wang (Renmin Univ, China), Kevin Wilkinson (HP), Yuhong Xiong (HP), Bin Zhang (HP), Bo Zhang (Tsinghua Univ, China), Ming Zhang (Peking Univ, China), Weimin Zheng (Tsinghua Univ, China)

Thesis Advisor and Postgraduate-Scholar Sponsor (5 years)
Post-doctoral scholars sponsored: Chetan Gupta (HP), Rui Liu (HP), Xiaomeng Huang (Tsinghua Univ.), Song Wang (HP)

HENRY D. JACOBY

Professor of Management
Sloan School of Management (Room E40-439)
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139
Email: Hjacoby@mit.edu
Phone: 617-253-6609, Fax: 617-253-9845

Professional Preparation
University of Texas at Austin, Mechanical Engineering, B.S.,1957.
Harvard University, Public Administration, MPA, 1963.
Harvard University, Economics, Ph.D., 1967.

Appointments
1991-   **Co-director, M.I.T. Joint Program on the Science and Policy of Global Change**
1990-2001 William F. Pounds Professor of Management, M.I.T
1989-91   Chair, M.I.T. Faculty
1980-83   Associate Director, M.I.T. Energy Laboratory
1978-80   Director, M.I.T. Center for Energy Policy Research
1973-   **Professor of Management, Massachusetts Institute of Technology**
1969-73   Associate Professor, John F. Kennedy School of Government, Harvard University
1965-69   Director, Harvard Water Program and Environmental Systems Program
1965-69   Instructor, and Assistant Professor, Department of Economics, Harvard University
1963-65   Harvard Development Advisory Service, Argentina Project, Buenos Aires
1959-61   Systems Analyst, Tudor Engineering Co., San Francisco, California
1957-59   Civil Engineering Corps, U.S. Navy

Publications Related to Current Project
Prinn, R., H. Jacoby, A. Sokolov, C. Wang, X. Xiao, Z. Yang, R. Eckaus, P. Stone, A. Ellerman, J.
   Melillo, J. Fitzmaurice, D. Kicklighter, G. Holian and Y. Liu (1999). Integrated Global System
   Model for Climate Policy Assessment: Feedbacks and Sensitivity Studies, *Climatic Change*
   41:469-456.
Reilly, J. P. Stone, C. Forest, M. Webster, H. Jacoby and R. Prinn (2001). Uncertainty and
   Climate Change Assessments, *Science* 293(5529): 430-33.
Forest, C., P. Stone and H. Jacoby (2002). How to Think About Human Influence on Climate,
   Forum for Applied Research and Public Policy 17(2): 47-51.
Jacoby, H. (2004). Modeling Human-Climate Interaction, S. Sparks and C. Hawksworth (eds.),
   The State of the Planet: Frontiers and Challenges in Geophysics, American Geophysical Union,
   Washington, D.C.
U.S. CCSP, 2007: U.S. Climate Change Science Program, Synthesis and Assessment Product
   2.1A: *Scenarios of Greenhouse Gas Emissions and Atmospheric Concentrations* (L. Clark,
   J. Edmonds, **H. Jacoby**, H. Pitcher, J. Reilly, R. Richels), in press
   (http://www.climatescience.gov/Library/sap/sap2-1/default.php).

Other Significant Publications
M. Babiker, J.M. Reilly and H.D. Jacoby (1999). The Kyoto Protocol and Developing Countries",
   Energy Policy 28:525-536.

Jacoby, H. (2004). Toward a Framework for Climate Benefits Estimation, in J. Corfee-Morlot and S. Agrawala (eds.), *The Benefits of Climate Change Policies: Analytical and Framework Issues*, OECD, Paris.

Schafer, A., and **H. Jacoby** (2005). Technology Detail in a Multi-Sector CGE Model: Transport Under Climate Policy, *Energy Economics* 27/1: 1-24.

Schafer, A., and **H. Jacoby** (2005). Vehicle Technology under CO_2 Constraint: A General Equilibrium Analysis, *Energy Policy* 34/9: 975-985.

Paltsev, S., **H. Jacoby**, J. Reilly, L. Viguier and M. Babiker (2005). Modeling the Transport Sector: The Role of Existing Fuel Taxes in Climate Policy, in R. Loulou et al. (eds.), *Energy and Environment*, Klewer Academic Publishers, Montreal.

Paltsev, S., J. Reilly, **H. Jacoby** & K.-H. Tay (2005). The Cost of Kyoto Protocol Targets: The Case of Japan, *Energy Economics*, in press.

Jacoby, H., J. Reilly, J. McFarland and S. Paltsev (2006). Technology and Technical Change in the MIT EPPA Model, *Energy Economics* 28/4: 610-631.

Schafer, A. and **H. Jacoby** (2006). Experiments with a Hybrid-MARKAL Model, *The Energy Journal*, Special Issue on Hybrid Modeling: 171-177.

Sekar, R., J. Parsons, H. Herzog and **H. Jacoby** (2007). Future carbon regulations and current investments in alternative coal-fired power plant technologies, *Energy Policy*, 35(2): 1064-1074.

Ansolobehere, S., J. Beer, J. Deutch, A.D. Ellerman, J. Friedman, H. Herzog, **H. Jacoby**, P. Joskow, G. McRae, R. Lester, E. Moniz and E. Steinfeld (2007). *The Future of Coal: Options for a Carbon-Constrained World*, an interdisciplinary MIT study, Cambridge, MA.

Paltsev, S., J. Reilly, **H. Jacoby**, and K. Tay, 2007: How (and why) do climate policy costs differ among countries? Chapter 24 in M.E. Schlesinger, H.S. Kheshgi, J. Smith, F.C. de la Chesnaye, J.M. Reilly, T. Wilson, and C. Kolstad (eds.), *Human-Induced Climate Change: An Interdisciplinary Assessment*, Cambridge University Press, Cambridge: 282-293.

Synergistic Activities

Member of the Scientific Committee of the International Geosphere-Biosphere Program, and member of the Climate Research Committee, U.S. National Research Council

Supervision of development of economic components of the MIT Integrated Global System Model of the human-climate system.

Collaborators and Other Affiliations

(i) Collaborators, last 48 months
   J. Melillo, Terrestrial Ecosystems Center, Marine Biology Laboratory
   G. Metcalf, Tufts University
   A. Schäfer, Cambridge University
   M. Zimmerman, Michigan University
(ii) Graduate and Post Doctoral Advisors
   R. Dorfman, Department of Economics, Harvard University
(ii) Thesis Advisor and Postgraduate-Scholar Advisor, last 5 years
   Andreas Schäfer, Cambridge University
   Ian Sue Wing, Boston University

   Total students supervised (including MS students): 15.

1 Biographical Sketches

1.1 Vita

David Karger (A.B. Summa cum laude in Computer Science, 1989, Harvard University; Certificate of Advanced Study in Mathematics, Cambridge University, 1990; Ph.D., 1994, in Computer Science, Stanford University) is a Professor of Computer Science and a member of the Computer Science and AI Laboratory at the Massachusetts Institute of Technology. He is a recipient of the ACM 1994 Doctoral Dissertation Award, the 1997 Tucker prize, and the 2003 National Academy of Science Award for Innovations in Research. He is the past recipient of a National Science Foundation CAREER award (1996), an Alfred P. Sloane Foundation Fellowship (1997) and a Packard Foundation Fellowship (1998). He held a National Science Foundation graduate fellowship from 1990-93 and a Hertz Foundation graduate fellowship from 1993-94.

Professor Karger's research interests include algorithms and information retrieval. His work in algorithms has focused on applications of randomization to graph optimization problems and led to significant breakthroughs on several core problems. He has pursued applications of his algorithms research, contributing to research on large-scale peer-to-peer systems and participating in the founding of Akamai Technologies. His work on information retrieval has focused on developing tools for flexible semistructured data management by end-users without specialized skills; he led the development of the Haystack personal information management framework and is one of the PIs on the MIT SIMILE project.

1.2 Related Publications

The 5 publications most closely related to the proposed research are

1. Huynh, David, Miller, Robert, and Karger, David R. Potluck: Data Mash-Up Tool for Casual Users. 6th *International Semantic Web Conference (ISWC)*, November 2007

2. Huynh, David, Miller, Robert, and Karger, David R. Exhibit: Lightweight Structured Data Publishing. *16th International World Wide Web Conference*, 2007

3. Karun Bakshi and David Karger. Personalized Semantic Web Application Development by End Users. *Proceedings of the 1st Workshop on The Semantic Desktop. 4th International Semantic Web Conference* Galway, Ireland. 2005.

4. Dennis Quan, David Huynh, David Karger, and Robert Miller. User interface continuations. In *Proceedings of UIST (User Interface Systems and Technolgies*, 2003.

5. Jaime Teevan, Christine Alvarado, Mark Ackerman, and David R. Karger. The perfect search engine is not enough: A study of orienteering behavior in directed search. In *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, 2004.

Another 5 publications are

1. Douglas Cutting, David R. Karger, Jan Pedersen, and John W. Tukey. Scatter/gather: A cluster-based approach to browsing large document collections. In *Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 318–329, Copenhagen, Denmark, 1992.

2. David R. Karger, Eric Lehman, Tom Leighton, Matt Levine, Daniel Lewin, and Rina Panigrahy. Consistent hashing and random trees: Distributed caching protocols for relieving hot spots on the world wide web. In *Proceedings of the* 29th *ACM Symposium on Theory of Computing*, pages 654–663. ACM, ACM Press, May 1997.

3. Jaime Teevan and David R. Karger. Empiricial development of an exponential probabilistic model for text retrieval: Using textual analysis to build a better model. In 26th *Internationl ACM SIGIR Conference*, Toronto, July 2003. ACM SIGIR, ACM.

4. Dennis Quan, David Huynh, and David R. Karger. Haystack: A platform for authoring end user semantic web applications. In 2nd *International Semantic Web Conference*, Sanibel Island, FL, October 2003.

5. Dennis Quan and David R. Karger. Haystack: A user interface for creating, browsing and organizing arbitrary semistructured information. In *ACM CHI Conference on Human Factors in Computing Systems*, 2004.

1.3 Collaborators and Other Affiliations

Collaborators in the past 48 months: Mark S. Ackerman, Christine Alvarado, Karun Bakshi, Hari Balakrishnan, Regina Barzilay, Michael Bernstein, Chris Bizer, Avrim Blum, Matthew Brand, A. Chan, Shuchi Chawla, Harr Chen, Austin T. Clements, Isaac G. Councill, Frank Dabek, Pawan Deshpande, Michelle Effros, Joan Feigenbaum, Jon Feldman, Nicholas J. A. Harvey, B. Hassibi, Joe Hellerstein, Tracey Ho, Andrew Hogue, David F. Huynh, Nicole Immorlica, Piotr Indyk, William Jones, M. Frans Kaashoek, Boris Katz, Max Van Kleek, Philip N. Klein, Ralf Koetter, Terran Lane, Ryan Lee, B. Leong, Jinyang Li, David Liben-Nowell, Jimmy Lin, Book Thau Loo, Muriel Médard, Raghu Madyastha, Stefano Mazzochi, Adam Meyerson, Robert C. Miller, Maria Minkoff, Vahab S. Mirrokni, Robert Morris, Kazuo Murota, Evdokia Nikolova, Emmanuel Pietriga, Dan R. K. Ports, Dennis Quan, S. Ray, Jason D. M. Rennie, Matthias Ruhl, Rahul Sami, m. c. schraefel, Yuan Kui Shen, Scott Shenker, J. Shi, Kai Shih, Lawrence Shih, Vineet Sinha, Clifford Stein, Ion Stoica, Jeremy Stribling, Jaime Teevan, Mikkel Thorup, Martin J. Wainwright, Michael Walfish, Sergey Yekhanin, Neal Young, J. D. Zamfirescu

Thesis Advisor and Postdoctoral Sponsor of: Matt Levine. Raj Iyer. Sudipta Sengupta. Maria Minkoff. Dennis Quan. Karun Bakshi. Vineet Sinha. Jaime Teevan. David Huynh. Kai Shih. Jon Feldman. Matthias Ruhl. Rahul Sami.

Advisors:

- Rajeev Motwani, Stanford University

- Serge Plotkin, Stanford University

- Andrew Goldberg, NEC

Michele Kimpton
Executive Director
DSpace Foundation
Massachusetts Institute of Technology
77 Massachusetts Ave. E25-131
Cambridge, MA 02139

PROFESSIONAL PREPARATION
Lehigh University, Bethlehem PA; BS, 1984
Santa Clara University, Santa Clara CA; MBA, 1989

APPOINTMENTS
1. Director of Web Archive, Internet Archive 2002-2006
2. Vice President, Business Development and Marketing, Eframes.com Corp. 1998 –2001
3. Raychem Corporation, Menlo Park, CA 1984-1998
4. WW Director of Channel Marketing, Polyswitch Division 1992-1998
5. Sales Manager, PolySwitch Division 1990-1992
6. Product Manager 1988-1990
7. Product Development Engineer 1984-1988

RELATED PUBLICATIONS
1. G. Mohr, M. Kimpton et al, **"Introduction to Heritrix: a web archival crawler"**, proceedings for 4th international web archiving workshop, 2004.
2. M. Stack, M. Kimpton et al, **"Full text search of web archive collections",** proceedings for 4th international web archiving workshop, 2006
3. M. Kimpton, J. Ubois, "**Web Archiving",** 201-212, 2006
4. S.M. Schneider, K. Foot, M. Kimpton, G. Jones, **"Building Thematic web collections, Challenges and Experiences"**, - 3rd ECDL Workshop on Web Archives, Trondheim, Norway, August, 2003

SYNERGISTIC ACTIVITIES
1. Founding member of the International Internet Preservation C onsortium (IIPC), July 2003
2. Steering Committee for JA-SIG opensource software conference, Spring 2008
3. Steering C ommittee for SPARC open access digital repository conference, Fall 2008

RECENT COLLABORATORS (48 months)
Brewster Kahle – Internet Archive
Bill Arms – Cornell University
Raymie Stata - Yahoo

Isaac S. Kohane
Associate Professor in Pediatrics
Harvard Medical School
25 Shattuck Street
Boston, MA 02115

PROFESSIONAL PREPARATION
Brown University, Providence, RI; ScB, 1981
Brown University, Providence, RI; M.D, Medicine, 1987
Boston University, Boston, MA; Ph.D, Computer Science, 1987

APPOINTMENTS
1. Director, Countway Library of Medicine, Harvard Medical School, 2005
2. Director, Children's Hospital Informatics Program, 1995
3. Assistant in Medicine, Children's Hospital, Boston, MA, 1992
4. Senior Assistant Resident in Pediatrics, Children's Hospital, Boston, 1989-1990
5. Assistant Resident in Pediatrics, Children's Hospital, Boston, 1988-89

RELATED PUBLICATIONS
1. Butte AJ, Kohane IS, **"Creation and implications of a phenome-genome network"**, Nat Biotechnol. 2006, Jan;24(1):55-62. 1.
2. Inaoka, H., Y. Fukuoka, I.S. Kohane, "**Lower expression of genes near microRNA in C. elegans germline",** BMC Bioinformatics, 2006. 7(1): p. 112.
3. Lee, S., I. Kohane, S. Kasif, "**Genes involved in complex adaptive processes tend to have highly conserved upstream regions in mammalian genomes**", BMC Genomics, 2005. 6: p. 168.
4. Kohane, I.S., D.R. Masys, R.B. Altman, **"The incidentalome: a threat to genomic medicine. Jama",** 2006. 296(2): p. 212-5.
5. Liu, H., et al., **"Predicting Survival within the Lung Cancer Histopathological Hierarchy Using a Multi-Scale Genomic Model of Development"**. PLoS Med, 2006. 3(7): p. e232.
6. Carter, S.L., et al., A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. Nat Genet, 2006. 38(9): p. 1043-8.

SELECTED PUBLICATIONS
1. Sebastiani P, Lazarus R, Weiss ST, Kunkel LM, Kohane IS, Ramoni MF. "**Minimal haplotype tagging",** Proc Natl Acad Sci U S A 2003;100(17):9900-5.
2. Nimgaonkar A, Sanoudou D, Butte AJ, Haslett JN, Kunkel LM, Beggs AH, et al. "**Reproducibility of gene expression across generations of Affymetrix microarrays"** BMC Bioinformatics 2003;4(1):27.
3. Fukuoka Y, Inaoka H, Kohane IS. "**Inter-species differences of co-expression of neighboring genes in eukaryotic genomes**", BMC Genomics 2004;5(1):4.
4. Kohane IS, Altman RB. "**Health-information altruists--a potentially critical resource**" N. Engl J Med. 2005 Nov 10;353(19):2074-7.
5. Lee S, Kohane I, Kasif S. "**Genes involved in complex adaptive processes tend to have highly conserved upstream regions in mammalian genomes**", BMC Genomics. 2005 Nov 27;6:168.

SYNERGISTIC ACTIVITIES
1. NIH Grant # 1P01NS047572-01A1   (Stiles) 5/1/05-4/30/09
Molecular Mechanism of Fate Choice in Neural Stem Cells
Goal: Identify regulation of genes involved in glial cell development.
2. Grant: 1 TO1DK60837-01A1 (PI: Kahn) 09/01/02-7/31/08
Diabetes Genomic Initiative Project: bioengineering Research Partnership
Role: Co-Investigator
Goal: Provide web-accessible annotation, cataloging facilities and state-of-the art bioinformatics analyses.
3. Grant# N01-LM-3-3515 (PI: Kohane) 9/30/03-9/29/07

Ensemble for Self-Scaling Systems for Health
Goal: Integration of hospital information systems, emergency department information systems, personal health information systems on a regional scale to allow surveillance and clinical management in cases of public health disaster.
4. NIH Grant # 1RO1NS047527-01A1 (PI:Rowitch)   7/1/04-6/30/09
Mechanisms of Hedge-Hog Induced Neuroproliferation
Goal: To identify developmental-stage specific gene expression programs.
5. Grant # U54 LM008748 (PI: Kohane) 9/15/04-7/31/09
National Centers for Biomedical Computing
Informatics for Integrating Biology and the Bedside
Goal: Clinical Research in the genomic era.
6. Genetic Studies of Autism Spectrum Disorders 9/1/04-8/31/07
Nancy Lurie Marks Foundation
Goal: To identify genes and gene pathways that lead to the ASD phenotype and to subdivide the patients into diagnostic categories based on their gene expression profiles in peripheral blood.
7. Grant # P01 CD000260 (PI: Platt, Richard) 09/30/05-09/29/08
Enhancing Public Health Through Electronic Medical and Personal Health Records
Goal: The goal of the Center of Excellence in Public Health Informatics is to leverage electronic health records and personal health records for public health.
8. CDC Grant #1 R01 PH000040-01 (Reis/Mandl) 9/30/05-9/29/08
Biosense Project Scaling Biosense: Advance Informatics Solution for Critical Problems
Goal: The major goals of this study are to develop informatics solutions to critical problems in surveillance.

RECENT COLLABORATORS (48 months)
Kunkel, Louis Harvard Medical School, Alal Eran MIT, Kenneth Mandl, Children's Hospital, Kho, Alvin T, HMS, Chueh, Henry C. HMS, Ramoni, Marco F. HMS, Murphy, Shawn N., HMS, Kucherlapati, Raju HMS, Drazen, Jeffrey M. HMS, Altman, Russ B. Stanford University, Halamka, J. HMS, Kim, Tae-Kyung HMS, Williams, Gordon H. HMS, Beggs, Alan H.   HMS, Greenberg, Michael E. HMS, Sun, Yao, UCSF, Brownstein, John S. HMS, Harris, Lyndsay N. Yale University, Turchin, Alexander HMS, Haab, Brian B. HMS, Mariani, Thomas J. HMS, Murphy, George F. HMS, Sebastiani, Paola Boston University, Szallasi, Zoltan HMS, Adler, Gail K. HMS, Butte, Atul J. Stanford University, Szolovits, Peter MIT, Tschumperlin, Daniel J, HMS, Fukuoka, Yutaka Tokyo U. Medical School, Gussoni, Emanuela HMS, Kang, Peter B HMS, Riva, Alberto U. of Florida, Sanoudou, Despina University of Athens, Greece.

THOMAS W. MALONE

BIOGRAPHICAL SKETCH

Education:
Rice University, Mathematical Sciences, B.A. (magna cum laude), 1974
Stanford University, Psychology, M.A., 1977
Stanford University, Engineering-economic systems, M.S., 1979
Stanford University, Psychology, Ph.D., 1980

Appointments and Professional Experience:
MIT Sloan School of Management (Cambridge, MA)

2006-present	Director, MIT Center for Collective Intelligence
2004-present	Patrick J. McGovern Professor of Management
2000-present	Head, Information Technology Group
1989-2006	Director, Center for Coordination Science
1989-2004	Patrick J. McGovern Professor of Information Systems
1994-1999	Co-Director, Initiative on Inventing the Organizations of the 21st Century
1985-1989	Associate Professor
1983-1985	Assistant Professor

IESE Business School (Barcelona, Spain)

2001-2002	Visiting Professor

Harvard Business School (Boston, MA)

1992	Visiting Professor

Xerox Palo Alto Research Center (Palo Alto, CA)

1980-1983	Member of Research Staff
1979-1980	Research Intern

Region IV Education Service Center (Houston, TX)

1974-1975	Consultant for Computer-Based Instruction

Publications – five most closely related to proposed project:
Malone, T. W., *The Future of Work: How the New Order of Business Will Shape Your Organization, Your Management Style, and Your Life*. Boston, MA: Harvard Business School Press, 2004 (Translated into Japanese, Chinese, Spanish, Korean, Portuguese, and Russian).
Malone, T. W. Bringing the market inside. *Harvard Business Review*, April 2004, *82* (4), 106-114.
Malone, T. W., Crowston, K. G., & Herman, G. (Eds.), *Organizing Business Knowledge: The MIT Process Handbook*. Cambridge, MA: MIT Press, 2003.
Malone, T. W., Laubacher, R. J., & Scott Morton, M. S. (Eds.), *Inventing the Organizations of the 21st Century*. Cambridge, MA: MIT Press, 2003.
Malone, T. W. & Crowston, K. The interdisciplinary study of coordination. *ACM Computing Surveys*, 1994 (March), 26 (1), 87-119.

Publications – five other significant:
Olson, G. M., Malone, T. W., and Smith, J. B. (Eds.), *Coordination Theory and Collaboration Technology*. Mahwah, NJ: Erlbaum, 2001.
Malone, T. W., Crowston, K. G., Lee, J., Pentland, B., Dellarocas, C., Wyner, G., Quimby, J., Osborn, C. S., Bernstein, A., Herman, G., Klein, M., & O'Donnell, E., "Tools for inventing organizations: Toward a handbook of organizational processes". *Management Science*, 1999, 45, 3 (March), 425-443. (Reprinted in several books.)

Malone, T. W. & Laubacher, R. J., "The Dawn of the E-lance Economy". *Harvard Business Review*, Sep. – Oct. 1998, 76 (5), 144-152. (Original and revised version each reprinted in several books.)

Malone, T.W., Crowston, K., Lee, J., Pentland, B., & Dellarocas, C. Computer system for displaying representations of processes. US Patent No. 5,819,270 (October 6, 1998).

Malone, T. W., Yates, J., & Benjamin, R. I. "Electronic markets and electronic hierarchies", *Communications of the ACM*, 1987, 30, 484-497. (Reprinted in two books.)

Synergistic Activities:

Developed innovative MIT course on "Strategic Organizational Design" incorporating traditional organizational design theory, concepts of collective intelligence, and numerous examples of other innovative new organizational designs.

Principal investigator and project leader for MIT Process Handbook project. Project resulted in creation of extensive, publicly available, on-line repository of knowledge about over 5000 business processes and activities. (http://ccs.mit.edu/ph)

Co-founder and chairman of Phios Corp., MIT spin-off company created to commercialize the results of the MIT Process Handbook Project.

Program chair, Conference on Computer-Supported Cooperative Work (1994).

Congressional testimony: Testified for US House of Representatives Committee on Education and the Workforce (Subcommittee on Oversight and Investigations), American Worker at a Crossroads Project, October 29, 1997.

Collaborators (last 48 months) and Co-Editors (last 24 months):

Current affiliation is MIT unless otherwise indicated: Deborah Ancona, Tom Apel, Michael Barrett (Cambridge U.), Abraham Bernstein (U. Zurich), Erik Brynjolfsson, Chrysanthos Dellarocas (U. Md.), Kevin Crowston (Syracuse), Victoria D'Urso (U. of Tennessee), Peter Denning (Naval Postgraduate School), Paulo Goncalves (University of Miami), Benjamin Grosof, George Herman, James Hines (Ventana Systems), Hiroshi Ishii, Mark Klein, S. P. Kothari, Richard Lai (Harvard), Robert Laubacher, Jintae Lee (U. Colorado), David McAdams, Michael Scott Morton, Mary Murphy-Hoye (Intel), Elissa O'Donnell (Phios), Wanda Orlikowski, James Patten (Patten Studio), John Quimby, James Rice, Peter Senge, Brian Subirana (IESE), Peter Weill,, Stephanie Woerner, JoAnne Yates.

Graduate Advisors (supervisors):

Patrick Suppes (Stanford); Mark Lepper (Stanford); John Seely Brown (USC), Robert Calfee (Stanford, UC Riverside).

Doctoral thesis advisees (primary advisor):

Abraham Bernstein (U. Zurich), Chrysanthos Dellarocas (U. Md.), George Wyner (Boston U.), Kevin Crowston (Syracuse), Paul Resnick (U. Michigan), Mark Ackerman (U. Michigan), David Rosenblitt, Erik Brynjolfsson (MIT)

Doctoral thesis advisees (committee member):

Brian Pentland (Michigan State), Lorin Hitt (U. Penn.), Albert Wenger, Jintae Lee (U. Colorado), Barry Arons, Kris Thorisson, Regina Ferreira Bento, Wendy Mackay, David Murotake, Andrew Trice, David Robertson

Number of Doctoral Theses advised as primary advisor (career): 8
Number of Postdoctoral scholars advised (career): 2

BIOGRAPHICAL SKETCH

Provide the following information for the key personnel and other significant contributors in the order listed on Form Page 2.
Follow this format for each person. **DO NOT EXCEED FOUR PAGES.**

NAME McCray, Alexa T.	POSITION TITLE Associate Professor of Medicine, Harvard Medical
eRA COMMONS USER NAME MCCRAY	School

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, and include postdoctoral training.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	YEAR(s)	FIELD OF STUDY
Skidmore College, Saratoga Springs, NY	BA	1969	Modern Languages
Boston College, Boston, MA	MA	1972	German Lit. & Language
Georgetown University, Washington, DC	MS	1977	Linguistics
Georgetown University, Washington, DC	PhD	1981	Linguistics, w/ distinction

A. Positions and Honors.

Positions and Employment

1979-1980 Research Affiliate, Linguistics & Philosophy, MIT, Cambridge, MA
1981-1984 Assistant Professor, Linguistics, Georgetown University, Washington, DC
1984-1985 Research Staff Member, Computer Sciences, IBM T.J. Watson Research Center, Yorktown,NY
1986-1992 Scientist, Lister Hill National Center for Biomedical Communications, NLM, Bethesda, MD
1992-1997 Branch Chief, Lister Hill National Center for Biomedical Communications, NLM, Bethesda, MD
1997-2004 Director, Lister Hill National Center for Biomedical Communications, NLM, Bethesda, MD
2005-2006 Lecturer on Medicine, Harvard Medical School
2005-present Co-Director, Center for Biomedical Informatics; Deputy Director, Countway Library of Medicine, Harvard Medical School
2006-present Associate Professor of Medicine, Harvard Medical School

Other Experience and Professional Memberships

Member of the Executive Committee, The Autism Consortium, 2006- ; Member of the Board, American Medical Informatics Association, 1999-2000; Member of the Board, International Medical Informatics Association, 1997-1999; Co-Editor-in-Chief, *Methods of Information in Medicine*, 2001- ; Editorial Board Member, *Journal of the American Medical Informatics Association*, 1995- ; Editorial Board Member, *Methods of Information in Medicine,* 1999-2001; Faculty, *Medical Informatics: A course for health professionals*, October 2000- ; Member, U.S. Government Senior Executive Service, 1997-2004; Chair, Informatics Working Group, The Autism Consortium, 2006- ; Chair, Publications Committee, American Medical Informatics Association, 1998-2001; Member, Publications Committee, American Medical Informatics Association, 1994-1997; Member, Executive Committee, American College of Medical Informatics, 1995-1999; Member, program committees Medinfo 2004, ISMB/ECCB 2004, Joint Conference on Digital Libraries, 2001-2004, ACM Digital Libraries, 1998-2000, IEEE Advanced Digital Libraries, 1996-1999, American Medical Informatics Association, Fall Symposium, 1996, 1997, 2003, 2007; Chair, Scientific Program Committee, Medinfo 2007.

Honors

Member, Institute of Medicine of the National Academy of Sciences, elected 2001
Fellow, American Association for the Advancement of Science, elected 1998
Fellow, American College of Medical Informatics, elected 1994
Innovations in American Government Award, for ClinicalTrials.gov, Harvard University, presented 2004
Priscilla Mayden Award, American Medical Informatics Association, presented 1996
Vice President Gore's Plain Language Award for ClinicalTrials.gov, presented 2001
NLM Director's Award, presented 2000
NIH Director's Award, presented 1997

B. Selected peer-reviewed publications (in chronological order).

(Publications selected from 60 peer-reviewed publications)

1. McCray AT, Srinivasan S. Automated access to a large medical dictionary: online assistance for research and application in natural language processing. Comput Biomed Res. 1990; 23(2):179-98.
2. Lindberg DA, Humphreys BL, McCray AT. The Unified Medical Language System. Methods Inf Med. 1993; 32(4):281-91.
3. McCray AT, Aronson AR, Browne AC, Rindflesch TC, Razi A, Srinivasan S. UMLS knowledge for biomedical language processing. Bull Med Libr Assoc. 1993; 81(2):184-94.
4. McCray AT. Representing biomedical knowledge in the UMLS Semantic Network. High-Performance Medical Libraries: Advances in Information Management for the Virtual Era.   Westport: Meckler Publishing, 1993; 45-55.
5. McCray AT, Nelson SJ. The representation of meaning in the UMLS. Methods Inf Med. 1995; 34(1-2):193-201.
6. Humphreys BL, Hole WT, McCray AT, Fitzmaurice JM. Planned NLM/AHCPR large-scale vocabulary test: using UMLS technology to determine the extent to which controlled vocabularies cover terminology needed for health care and public health. J Am Med Inform Assoc. 1996; 3(4):281-7.
7. McCray AT Conceptual complexity in biomedical terminologies: The UMLS approach. Classification and Knowledge Organization, Berlin: Springer Verlag, 1997; 475-89.
8. Humphreys BL, McCray AT, Cheh ML. Evaluating the coverage of controlled health data terminologies: report on the results of the NLM/AHCPR large scale vocabulary test. J Am Med Inform Assoc. 1997; 4(6):484-500.
9. McCray AT. The nature of lexical knowledge. Methods Inf Med. 1998; 37(4-5):353-60.
10. McCray AT. A national resource for information on clinical trials. National Forum. 1999; 17:19-21.
11. McCray AT, Ide NC. Design and implementation of a national clinical trials registry. J Am Med Inform Assoc. 2000; 7(3):313-23.
12. McCray AT. Better access to information about clinical trials. Ann Intern Med. 2000; 133(8):609-14.
13. McCray AT. Digital library research and application. Stud Health Technol Inform. 2000; 76:51-62.
14. McCray AT. Improving access to healthcare information: the Lister Hill National Center for Biomedical Communications. MD Comput. 2000; 17(2):29-34.
15. McCray AT, Gallagher ME. Principles for digital library development. Communications of the ACM 2001; 44(5):49-54.
16. Lederberg J, McCray AT. 'Ome sweet 'omics – A genealogical treasury of words. The Scientist, April 2, 2001; 8.
17. McCray AT. An upper level ontology for the biomedical domain. Comp Funct Genom 2003; 4:80-4.
18. Ammenwerth E, Wolff AC, Knaup P, Ulmer H, Skonetzki S, van Bemmel JH, McCray AT, Haux R, Kulikowski C. Developing and evaluating criteria to help reviewers of biomedical informatics manuscripts: A case report. J Am Med Inform Assoc. 2003;10:512-4.
19. Mitchell JA, McCray AT, Bodenreider O. From phenotype to genotype: Issues in navigating the available information resources. Methods Inf Med 2003; 42:557-63.
20. Bodenreider O, McCray AT. Exploring semantic groups through visual approaches. Journal of Biomedical Informatics 2003;36(6):414-432.
21. Mitchell JA, Fun J, McCray AT. Design of Genetics Home Reference: A new NLM consumer health resource. J Am Med Inform Assoc 2004;11(6):439-447.
22. Lomax J, McCray AT. Mapping the Gene Ontology into the Unified Medical Language System. Comp Funct Genom 2004; 354-61.
23. McCray AT. Promoting Health Literacy. J Am Med Inform Assoc 2005;12:152-163.
24. McCray AT. Conceptualizing the world: Lessons from history. J Biomed Inform 2006; 39(3):267-73.
25. Fomous C, Mitchell JA, McCray AT. Genetics Home Reference: helping patients understand the role of genetics in health and disease. Community Genetics 2006;9:247-8.

C. Research Support
Federal employee from January 1986 – January 2005
5283138-01 (PI:McCray)   3/1/2007 – 2/28/2008
Autism Consortium, LLC
Integration and Standardization of Autism Consortium Data
The major goal of this project is to ensure that Autism Consortium research efforts constitute a sustainable, collaborative research effort across phenotypes and multiple measurement modalities for lasting use by the Consortium and the greater developmental medicine research community.

Jill P. Mesirov

Professional Preparation

University of Pennsylvania	Mathematics	A.B.	1970
Brandeis University	Mathematics	M.A.	1971
Brandeis University	Mathematics	Ph.D.	1974

Appointments

11/2003-Present	**The Eli and Edythe L. Broad Institute, Massachusetts Institute of Technology and Harvard University,** Associate Director and Chief Informatics Officer, Director, Bioinformatics and Computational Biology
6/1997-11/2003	**Whitehead Institute Center for Genome Research,** Associate Director and Chief Informatics Officer, Director, Bioinformatics and Computational Biology Program
2001-Present	**Boston University,** Adjunct Professor of Bioinformatics
1995-1997	**International Business Machines Corporation,** Manager, Computational Biology and Bioinformatics
1985-1995	**Thinking Machines Corporation,** Director of Research, Senior Scientist
1983-1987	**1986 International Congress of Mathematicians,** Executive Director
1982-1985	**American Mathematical Society,** Associate Executive Director
1980 and 1982	**Princeton University** Visiting Lecturer
1978	**Australian National University,** Visiting Fellow
1976-1982	**Institute of Defense Analyses,** Research Staff
1974-1976	**University of California at Berkeley,** Instructor, Department of Mathematics

Five relevant papers:

1. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, **Mesirov JP**, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science. 1999 Oct 15;286(5439):531-7.
2. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, Puigserver P, Carlsson E, Ridderstrale M, Laurila E, Houstis N, Daly MJ, Patterson N, **Mesirov JP**, Golub TR, Tamayo P, Spiegelman B, Lander ES, Hirschhorn JN, Altshuler D, Groop LC. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet. 2003 Jul;34(3):267-73.
3. Brunet J-P, Tamayo P, Golub T, **Mesirov JP**. Metagenes and Molecular Pattern Discovery Using Matrix Factorization, Proc. Natl. Acad. Sci. USA. 2004;101(12):4164-4169.
4. Monti S, Tamayo P, **Mesirov JP**, Golub T. Consensus Clustering. A resampling-based method for class discovery and visualization of gene-expression microarray data. Functional Genomics special issue, Machine Learning. 2003;52:91-118.
5. Subramanian A, Tamayo P, Mootha V, Mukherjee S, Ebert BL, Gillette MA, Pomeroy S, Golub TR, Lander ES, **Mesirov JP**. Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-wide Expression Profiles. (submitted, Proc. Natl. Acad. Sci.).

Five other papers:

1. Sweet-Cordero A, Mukherjee S, Subramanian A, You H, Roix J, Ladd C, **Mesirov JP**, Golub T, Jacks T. An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis. Nature Genetics. 2005 Jan;37(1):48-55.

2. Reich M, Ohm K, Tamayo P, **Mesirov JP**. GeneCluster 2: advanced tools for bioarray analysis. Bioinformatics. 2004 Jul 22;20(11):1797-8.
3. Liefeld T, Reich M, Lerner J, Gould J, Zhang P, Tamayo P, **Mesirov JP**. GeneCruiser: A web service for the annotation of microarray data. (in press, Bioinformatics).
4. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature. 2001 Feb 15;409(6822):860-921. Erratum in: Nature 2001 Aug 2;412(6846):565. Nature 2001 Jun 7;411(6838):720.
5. Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. Nature. 2002 Dec 5;420(6915):520-62.

Synergistic Activities (*selected current activities*)
    Department of Energy: Argonne National Laboratory, MCS Review Committee
    International Society for Computational Biology: Board of Directors
    Los Alamos National Laboratory: Bioscience Division Review Committee
    Pittsburgh Super Computer Center: NIH Research Resource Advisory Committee
    Industrial Scientific Advisory Boards: Bristol-Myers Squibb, Infinity Pharmaceuticals Inc., Charles River Laboratory

Collaborators & Other Affiliations
    (a) Collaborators and Co-Editors
        Bonnie Berger, Massachusetts Institute of Technology
        Serafim Batzoglou, Stanford University
        Andrea Califano, First Genetic Trust
        Leif C. Groop, Lund University, Sweden
        Cecilia M. Lindgren, Lund University, Sweden
        Hongjun Lu, Hong Kong University of Science and Technology
        Dimitris Meretakis, Hong Kong University of Science and Technology
        Lior Pachter, University of California, Berkeley
        Tomaso Poggio, Massachusetts Institute of Technology
        Scott Pomeroy, Harvard Medical School
        Pere Puigserver, Dana-Farber Cancer Institute
        Ryan Rifkin, Massachusetts Institute of Technology
        Margaret Shipp, Dana-Farber Cancer Institute
        Donna Slonim, Genetics Institute
        Bruce Spiegelman, Dana-Farber Cancer Institute
        Gustavo Stolovitsky, IBM
        John Weinstein, National Cancer Institute
        Beat Wüthrich, Hong Kong University of Science and Technology
        Chen-Hsiang Yeang, Massachusetts Institute of Technology

    (b) Graduate and Post Doctoral Advisors
        Richard S. Palais, Brandeis University

    (c) Thesis Advisor and Postgraduate-Scholar Sponsor
        Pablo Tamayo   Xiru Zhang
        Serafim Batzoglou   Total number of graduate students – 6
        Jean-Philippe Brunet   Total number of post-docs – 3
        Elinor Karlsson
        Tarjei Mikkelsen
        Sayan Mukherjee
        Cynthia Phillips
        Aravind Subramanian

RONALD G. PRINN

TEPCO Professor of Atmospheric Sciences, Department of Earth, Atmospheric and Planetary Sciences
Massachusetts Institute of Technology, Room 54-1312, 77 Massachusetts Ave., Cambridge, MA 02139
Phone: 617-253-2452, Email: rprinn@mit.edu, Website: http://mit.edu/rprinn/

Professional Preparation

University of Auckland, New Zealand, Chemistry, and Pure and Applied Mathematics, B.Sc., 1967
University of Auckland, Chemistry (with first class honors), M.Sc., 1968
Massachusetts Institute of Technology, Chemistry, Sc.D., 1971

Appointments

1993-present: *TEPCO Professor* of Atmospheric Sciences, MIT
1998-2003: *Head*, Department of Earth, Atmospheric, and Planetary Sciences, MIT
1991-present: *Co-Director*, MIT Joint Program on the Science and Policy of Global Change
1990-present: *Director*, MIT Center for Global Change Science
1981: *Visiting Associate Professor*, California Institute of Tech., Div of Geological & Planetary Sciences
1971-1992: *Professor* (*Assistant*, 1971; *Associate*, 1976; *Full*, 1982), MIT

Five Publications Closely Related to this Proposal (http://web.mit.edu/rprinn/)

Xiao, X., **R. G. Prinn**, P.G. Simmonds, L.P. Steele, P.C. Novelli, J. Huang, R.L. Langenfelds, S. O'Doherty, P.B. Krummel, P.J. Fraser, L.W. Porter, R.F. Weiss, P. Salameh, and R.H.J. Wang, Optimal estimation of the soil uptake rate of molecular hydrogen from AGAGE and other measurements, *J. Geophys. Res.*, **112,** D07303, doi:10.1029/2006JD007241, 2007.

Chen, Y.-H. and **R.G. Prinn**, Estimation of atmospheric methane emissions between 1996 and 2001 using a 3D global chemical transport model. *J. Geophys. Res.,* **111**, D10307, doi: 10.1029/2005JD006058, 2006.

Prinn, R.G., J. Huang, R.F. Weiss, D.M. Cunnold, P.J. Fraser, P.G. Simmonds, A. McCulloch, C. Harth, S. Reimann, P. Salameh, S. O'Doherty, R.H.J. Wang, L. Porter, B.R. Miller and P. Krummel, Evidence for variability of atmospheric hydroxyl radicals over the past quarter century, *Geophys. Res. Lttr.*, **32**, L07809, doi: 10.1029/2004GL022228, 2005.

Prinn, R. and S. Dorling, Climate change and air quality: International perspectives and policy implications, *Environmental Management*, 37-40, October, 2005.

Webster, M., C. Forest, J. Reilly, M. Babiker, D. Kicklighter, M. Mayer, **R.G. Prinn**, M. Sarofim, A. Sokolov, P. Stone and C. Wang, Uncertainty Analysis of Climate Change and Policy Response, *Climatic Change*, **61**, 295-320, 2003.

Five Other Recent Publications

Reilly, J., M. Sarofim, S. Paltsev, and **R. Prinn**, The role of non-CO_2 GHGs in climate policy: Analysis using the MIT IGSM, *Energy Journal*, Multi-greenhouse Gas Mitigation and Climate Policy Special Issue, 503-520, 2006.

Chen, Y.-H. and **R.G. Prinn**, Atmospheric modeling of high-frequency methane observations: Importance of interannually varying transport. *J. Geophys. Res.,* **110**, D10303, doi: 10.1029/2004JD005542, 2005.

Prinn, R.G., Non-CO_2 Greenhouse Gases, in *The Global Carbon Cycle*, ed. C. Field and M. Raupach, Island Press, Washington, D.C., pgs. 205-216, 2004.

Prinn, R.G., J. Huang, R.F. Weiss, D.M. Cunnold, P.J. Fraser, P.G. Simmonds, A. McCulloch, C. Harth, P. Salameh, S. O'Doherty, R.H.J. Wang, L. Porter, and B.R. Miller, Evidence for substantial variations of atmospheric hydroxyl radicals in the past two decades, *Science*, **292**, 1882-1888, 2001

Prinn, R.G., R.F. Weiss, P.J. Fraser, P.G. Simmonds, D.M. Cunnold, F.N. Alyea, S. O'Doherty, P. Salameh, B.R. Miller, J. Huang, R.H.J. Wang, D.E. Hartley, C. Harth, L.P. Steele, G. Sturrock, P.M. Midgley, and A. McCulloch, A history of chemically and radiatively important gases in air deduced from ALE/GAGE/AGAGE. *J. Geophys. Res.*, **115**, 17751-17792, 2000.

Prinn, R.G., Measurement equation for trace chemicals in fluids and solution of its inverse, in *Inverse Methods in Global Biogeochemical Cycles*, ed. P. Kasibhatla *et al.*, *Geophysical Monographs*, No. **114**, American Geophysical Union, pgs. 3-18, 2000.

Synergistic Activities (*Five Examples*)

Principal Investigator and leader, Advanced Global Atmospheric Gases Experiment (AGAGE) and its predecessors (ALE, GAGE) in which the rates of increase of the concentrations of the trace gases involved in the greenhouse effect and ozone depletion have been measured continuously over the globe since 1978.

Pioneering the use of inverse methods, which use the above measurements and three-dimensional models to determine trace gas emissions and understand atmospheric chemical processes.

Developed with colleagues the first comprehensive global three-dimensional dynamical-chemical-radiative model of the ozone layer and applied it to elucidating the effects of supersonic aircraft on ozone.

Developed with colleagues a unique integrated global system model coupling models of economics, climate physics and chemistry, and terrestrial ecosystems, and applied it to assessment of uncertainty in climate predictions and analysis of climate policies.

Made significant contributions to the development of national and international scientific research programs in global change (International Global Atmospheric Chemistry Program, International Geosphere-Biosphere Program).

Collaborators and their affiliation (*past 48 months*) (*other than at MIT, or advisees/postdocs listed below*)

Buchwitz, M., U. Bremen, Germany
Burrows, J., U. Bremen, Germany
Claussen, M., PIK, Germany
Cunnold, D., Georgia Tech.
de Beek, R., U. Bremen, Germany
Derwent, R., UK Met. Office
Felzer, B.S., MBL, Woods Hole, MA
Frankenberg, C., U. Heidelberg, Germ.
Fraser, P.J., CSIRO, Australia
Golombek, A., IIBR, Israel
Greally, B., U. Bristol, UK
Harth, C., SIO, UCSD
Hill, M., U. Sunderland, UK
Kicklighter, D., MBL, Woods Hole, MA
Kleiman, G., NESCAUM, Boston

Krummel, P., CSIRO, Australia
Langenfelds, R., CSIRO, Australia
Manning, A., UK Met. Office
McCulloch, A., U. Bristol, UK
McGuire, A.D., U. Alaska, Fairbanks
Melillo, J., MBL, Woods Hole, MA
Miller, B., SIO, UCSD
O'Doherty, S., U. Bristol, UK
Platt, U., U. Heidelberg, Germany
Porter, L., Cape Grim Monitoring Station, Tasmania
Pszenny, A.P., Univ. NH, Durham
Rasch, P., UK Met. Office
Rastetter, E.B., MBL Woods Hole, MA

Reimann, S., EMPA, Switzerland
Reiner, D., Cambridge U., UK
Sabine, C., NOAA/PMEL, Seattle, WA
Salameh, P., SIO, UCSD
Simmonds, P., U. Bristol, UK
Steele, L., CSIRO, Australia
Steudler, P.A., MBL, Woods Hole, MA
Sturrock, G.A., CSIRO, Australia
Taylor, J.A., Australian Nat'l U.
Wagner, T., U. Heidelberg, Germany
Wang, R.H.J., GeorgiaTech
Webster, M., Univ. NC, Chapel Hill
Weiss, R.F., SIO, UCSD
Williams, M., U. Edinburgh, UK
Zhuang, Q., Purdue U., IN

Graduate Advisees (*in the past 5 years*) **and current affiliation**

Yu-Han Chen (U.S. Dept. of Defense), Jin Huang (MIT), Don Lucas (Frontier Research Center for Global Change, Yokohama, Japan), Arnico Panday (Princeton University), Stephanie Shaw (University of California, Berkeley), Donnan Steele (McKinsey & Co., Houston, TX).

Postdoctoral Researchers Supervised (*in the past 5 years*) **and current affiliation**

Andrea Castanho (MIT), Annica Ekman (Stockholm Univ.), Felix Ng (Univ. of Sheffield, UK), Jose Ortega-Alija (Sandia National Lab.), Madelene Ostwald (Göteborg Univ., Sweden), Qian Tan (University of Maryland), Rainer Volkamer (UCSD)

Total number of graduate students advised: ~40; postdoctoral scholars sponsored: ~31

Michael D. Siegel
Sloan School of Management, E53-323
Massachusetts Institute of Technology
Cambridge, MA 02139
msiegel@mit.edu

BIOGRAPHICAL SKETCH

PROFESSIONAL PREPARATION

Education

1989	Ph.D. in Computer Science, Boston University, Boston, MA.
1985	M.A. in Computer Science, Boston University, Boston, MA.
1980	M.S. in Engineering, University of Wisconsin-Madison.
1977	B.S. in Engineering, Trinity College, Hartford, CT.

APPOINTMENTS / EMPLOYMENT (current in **bold**)

1993-	**Principal Research Scientist, Massachusetts Institute of Technology.**
1998	Senior Lecturer, Sloan School of Management, Massachusetts Institute of Technology.
1989-1993	Research Scientist, Sloan School of Management, MIT, Cambridge, MA.
1987	Member Technical Staff, Knowledge-Based Systems Department, GTE Laboratories, Waltham, MA
1985	Computer Scientist, Computer Corporation of America (Xerox Advanced Information Technology), Cambridge, MA.
Research areas:	intelligent integration of information systems, data semantics, data standards, web-based information extraction and integration, global risk management, information security, financial applications, pre-conflict analysis, voting technology, digital health and state stability.

PUBLICATIONS

Most Closely Related Publications (5)

2004	Zhu, H., S. Madnick, M. Siegel et al, "Effective Data Integration in the Presence of Temporal Semantic Conflicts" Conference on Temporal Representation and Reasoning, July 2004.
2002	Madnick, S., M. Siegel. "Seizing the Opportunity: Exploiting Web Aggregation", *MISQ Executive*, Vol 1, No. 1, March 2002, pp. 35-46. [SWP #4351, CeB #144, CISL #01-13].
2001	A. Moulton, S. Madnick, and M. Siegel, "Cross-Organizational Data Quality and Semantic Integrity: Learning and Reasoning about Data Semantics with Context Interchange Mediation", *Proceedings of the Americans Conference on Information Systems* (AMCIS, Boston), August 2001 [SWP #4167, CISL #01-04].
2000	Bresson, Stephane, C. Goh, N. Levina, A. Shah, and M. Siegel, "Context Knowledge Representation and Reasoning in the Context Interchange System," *The International Journal of Artificial Intelligence*, Neutral Networks, and Complex Problem-Solving Technologies, Volume 12, Number 2, September 2000, pp. 165-180, [SWP #4133, CISL #00-04].
1999	Goh, Cheng, Stephane Bresson, Stuart Madnick, and Michael Siegel , "Context Interchange: New Features and Formalisms for the Intelligent Integration of Information," *Transactions on Information Systems*, July 1999.

Additional Recent / Relevant Publications (5)

2006   Nazli Choucri, Christi Electris, Daniel Goldsmith, Dinsha Mistree, Stuart Madnick, J. Bradley Morrison, Michael Siegel, Margaret Sweitzer-Hamilton, "Understanding & Modeling State Stability: Exploiting System Dynamics, " *System Dynamics Conference*, Cambridge , MA, January 2006

2002   Madnick, S., M. Hansen and M. Siegel. "Data Integration using Web Services", *Proceedings of the Second International Workshop on Data Integration over the Web* (DIWeb'02, Toronto, Canada), May 2002 [CISL #2002-14], also to be published in *Lecture Notes in Computer Science*, Springev-Varlag, New York.

2001   Moulton, Allen, S. Madnick, M. Siegel , "Knowledge Representation Architecture for Context Interchange Mediation: Fixed Income Securities Investment Examples" (with), *Proceedings of the First International Workshop on Electronic Business Hubs:* XML, Metadata, Ontologies, and Business Knowledge on the Web (WEBH; Munich, Germany), published by IEEE (ISBN 0-7695-1230-5, ISSN 1529-4188), September 2001, pp. 50-54 [SWP #4169, CISL #01-05].

1995   Goh, Cheng, Stuart Madnick, and Michael Siegel, "Context Interchange: Overcoming the Challenges of Large-Scale Interoperable Database Systems in a Dynamic Environment," In *Proceedings of the Third International Conference on Information and Knowledge Management*, Gaithersburg, MD.

1994   Sciore, Edward, Michael Siegel and Arnie Rosenthal, "Using Semantic Values to Facilitate Interoperability Among Heterogeneous Information Systems," *ACM Transactions on Database Systems*, Vol. 19, No. 2, pp 254-290.

SYNERGISTIC ACTIVITIES

1. Developed technology and transferred for use by DOD and private companies.
2. Director of Digital Health Special Interest Group at the Digital Business Center, MIT. Research areas include data integration, real-time data collection systems, process improvement and data mining.
3. Served as Project Manager and Chief Research Scientist for various projects and programs, such as: Context Interchange (COIN), PROductivity From Information Technology (PROFIT), Pro-Active INTelligence (PAINT). Sponsors have included NSF, DARPA, iARPA, DOD, and industry companies.
4. Managed large private sector integration efforts for the financial services industry
5. Integration of System Dynamics modeling and analysis of large systems with new software applications and data collection efforts.

COLLABORTATORS AND OTHER AFFILIATIONS

Collaborators (5 years)
S. Bressan (National University of Singapore), **N. Choucri** (MIT), **A. Firat** (Northeastern U), **B. Grosof** (MIT), **Y. Lee** (Northeastern), **F. Manola** (independent consultant), **D. Mistree** (MIT), **B. Morrison** (Brandeis), **A. Moulton** (MIT), **S. Madnick** (MIT), **D. Strong** (WPI), **R. Wang** (MIT), **H. Zhu** (Old Dominion University).

Graduate and Postdoctoral Advisors
Edward Sciore (Boston College), Sharon Salveter (University of Chicago)

Thesis Advisor and Postgraduate-Scolar Sponsor: (5 years)
Thesis advisor: Wee Horng Ang, Aykut Firat, M. Bilal Kaleem, Philip Lee. (Approx. 50 in total)

Biographical Sketch: Michael Stonebraker

MIT Computer Science and Artificial Intelligence Laboratory
32 Vassar St., room 32-G922
Cambridge, MA. 02139
Tel (603) 714-4451
Email: stonebraker@csail.mit.edu

Areas of Special Interest
Database systems, data warehouses, federated database systems, database systems for new application areas, operating systems

Professsional Preparation:
B.S.E.E. - Princeton University 1961
PhD Computer Information and Control Engineering - Univ. of Michigan 1971

Appointments
University of California, Berkeley – Asst. Professor 1971-1976,
Assoc. Professor 1976-1982,
Professor 1982-1994,
Professor of the Graduate School 1994-1999.
Massachusetts Institute of Technology – Senior lecturer 2001-2002
Adjunct Professor 2002-
Ingres Corporation – Founder and Chief Technology Officer - 1980-1992
Illustra Corporation – Founder and Chief Technology Officer - 1992-1996
Informix Corporation – Chief Technology Officer - 1996-2000
Cohera Corporation – Founder and Chief Technology Officer - 1997-2001
Required Technology, Inc. Chief Technology Officer 2001- 2002
Streambase, Inc. – Founder and Chief Technology Officer – 2003-
Vertica Systems, Inc. – Founder and Chief Technology Officer – 2005 –

Awards and Honors:
ACM Software System Award 1988
ACM SIGMOD Innovation Award 1992
National Academy of Engineering 1998
Computer Reseller News – Hall of Fame Member, 1999
IEEE John von Neumann Medal 2005

Relevant Publications (5)
Michael Stonebraker et. al., "Whither DBMS Architecture," Proc 2007 CIDR Conference, Asilomar, Ca., Jan 2007
Michael Stonebraker, et. al., "C-Store: A Column-oriented DBMS," Proc. 2005 VLDB Conference, Trondheim, Norway, Sept 2005.
Michael Stonebraker, Ugur Cetintemel, "One Size Fits All: An Idea Whose Time Has Come and Gone," Proc. 2005 ICDE Conference, Tokyo, Japan, April 2005.
Hari Balakrishnan, Magda Balazinska, Don Carvey, Ugur Cetintemel, Mitch Cherniack, Christian Convey. Eddie Galvez, Jon Salz, Michael Stonebraker, Richard Tibbetts, Stan Zdonik, "Retrospective on Aurora," VLDB Journal 13(4), p370-383 (2004)
Daniel Abadi, Don Carvey, Ugur Cetintemel, Mitch Cherniack, Christian Convey.Sangdon, Michael Stonebraker, Nesime Tatbul, Stan Zdonik, "Aurora: A New Model and Architecture

for Data Stream Processing," Proc. 2002 VLDB Conference, Hong Kong, China August 2002.

Chris Olston, Allison Woodruff, Alexander Aiken, Michael Chu, Vuk Ercegovac, Mark Lin, Mybrid Spalding, Michael Stonebraker, "DataSplash," SIGMOD 1998, p550-552.

Michael Stonebraker, Paul M. Aoki, Witold Litwin, Avi Pfeffer, Adam Sah, Jeff Sidell, Carl Staelin, Andrew Yu, "Mariposa: A Wide-Area Distributed Database System," VLDB Journal 5(1), p48-63 (1996).

A sample of other publications (5)

Michael Stonebraker, Jolly Chen, Nobuko Nathan, Caroline Paxson, Jiang Wu, "Tioga: Providing Data Management Support for Scientific Visualization Applications," VLDB 1993, p25-38.

Michael Stonebraker, "The Integration of Rule Systems and Database Systems," TKDE 4(5), p415-423 (1992).

Michael Stonebraker, Greg Kemnitz, "The Postgres Next Generation Database Management System," CACM 34(10), p78-92 (1991).

Margo I. Seltzer, Michael Stonebraker, "Transaction Support in Read Optimizied and Write Optimized File Systems," VLDB 1990, p174-185.

Michael Stonebraker, Eugene Wong, Peter Kreps, Gerald Held, "The Design and Implementation of INGRES," TODS 1(3), p189-222 (1976).

Synergistic Activities:

- Wrote several major public domain DBMS prototypes, including C-Store, Ingres and POSTGRESQL. The latter system is currently widely used as a teaching instrument in universities and forms the basic for several commercial DBMS products.
- Have founded the New England Database Symposium (NEDS) as a mechanism for Boston-area DBMS researchers to network and collaborate.
- Have produced 4 editions of the "Red Book" – an edited collection of papers that forms the core readings in many graduate database courses.
- Have organized (in conjunction with Jim Gray) four self assessments (roughly every 5 years) of the DBMS field by senior researchers.
- Have started (in conjunction with Jim Gray and David Dewitt) the Conference on Data Systems Research (CIDR) – a conference oriented toward practical research in data systems.

Collaborations: Betty Salzberg (Northeastern), Stan Zdonik (Brown), Mitch Cherniak (Brandeis), Joey Hellerstein (Berkeley), Joachim Hammer (Florida), Sam Madden (M.I.T.), Pat O'Neil (UMass/Boston)

PhD Advisor: Dr. Arch Naylor (Michigan) – now retired

Completed PhD students: (sample from more than 20 completed PhD students) Gerald Held (until recently Vp/ Eng'g of Oracle Corp.), Robert Epstein (founder and until recently VP/Eng'g of Sybase Corp.), Paula Hawthorn (most recently VP/Eng'g of Informix Corp.), Dale Skein (founder and CTO of Vitria Corp.), Mike Carey (Senior Architect at BEA), Margo Seltzer (Assoc Professor of Computer Science at Harvard), Anant Jhingran (Manager - Database Department of IBM Almaden Research Laboratory)

Tyler O. Walters
Georgia Institute of Technology
704 Cherry St.
Atlanta, GA 30332-0900
Phone: (404) 385-4489; Fax: (404) 894-6084; Email: tyler@gatech.edu

PROFESSIONAL PREPARATION
Northern Illinois University, DeKalb, IL; BA in History, 1985.
North Carolina State University, Raleigh, NC; MA in History and Archival Management, 1988.
University of Arizona, Tucson, AZ; MA in Library and Information Sciences, 1996.

APPOINTMENTS
1. Associate Director, Technology and Resource Services; Georgia Tech Library and Information Center, December 2002 – present.
2. Director, William R. Haselton Library and Knowledge Center, Institute of Paper Science and Technology, 1998-2002.
3. Assistant to the Associate Director for Public Services and Collections, Iowa State University Library, 1998.
4. Head, Special Collections Department and Associate Professor, Iowa State University Library, 1996-1998.
5. University Archivist, Special Collections Department and Assistant Professor, Iowa State University Library, 1992-1996
6. Assistant University Archivist, University Archives, Northwestern University Library, 1988-1992

RELATED PUBLICATIONS
1. R. McDonald and T. Walters, "Sustainability Models for Digital Preservation Federations." Proceedings of DigCCurr 2007: An International Symposium in Digital Curation, 2007 http://www.ils.unc.edu/digccurr2007/papers/McDonaldWalters_paper_6-4.pdf
2. T. Walters, "The New Academic Library — Building Institutional Repositories to Support Changing Scholarly and Research Processes." Sailing into the Future – Charting our Destiny: Proceedings of the Association of College and Research Libraries 13th National Conference, 2007: 56-63
3. T. Walters, "Reinventing the Library: How Repositories Are Causing Librarians to Rethink Their Professional Roles," portal: Libraries and the Academy, v.7 April 2007: 213-225 http://hdl.handle.net/1853/14421
4. T. Walters, "Strategies and Frameworks for Institutional Repositories and the New Support Infrastructure for Scholarly Communications," D-Lib Magazine, October 2006 (http://www.dlib.org/dlib/october06/walters/10walters.html)
5. C. Thomas, R. McDonald, A. Smith, T. Walters, "The New Frontier of Institutional Repositories: A Common Destination with Different Paths," New Review of Information Networking, v.11 May 2005: 65-82.

SIGNIFICANT PUBLICATIONS
1. R. Meyer and T. Walters, "Sowing an Old Field with a New Crop: Collaborative Services of Libraries and Other Campus Units." Convergence and Collaboration of Campus Information Services. (Westport, CT: Libraries Unlimited) Peter Hernon and Ronald Powell, editors. Due for publication in 2008.

2. T. Walters, "Digital Sustainability: Weaving a Tapestry of Interdependency to Advance Digital Programs." Strategies for Sustaining Digital Libraries. (Atlanta: Emory University) Martin Halbert, editor. Due for publication in 2008.

3. P. De Stefano and T. Walters, "A Natural Collaboration: Preservation for Archival Collections in ARL Libraries." Library Trends: special issue: *Recent Trends in Cultural Heritage Preservation*. 56 (Summer 2007): 230-258.

SYNERGISTIC ACTIVITIES

1. Co-Principal Investigator, MetaArchive Cooperative, a founding partnership in the Library of Congress' National Digital Information Infrastructure and Preservation Program (NDIIPP), 2004 – present. http://www.metaarchive.org and http://www.digitalpreservation.gov

2. Invited participant in the Association of Research Libraries (ARL) workshop (2006), New Collaborative Relationships: The Role of Academic Libraries in the Digital Data Universe." The meeting produced the report, "To Stand the Test of Time: Long-term Stewardship of Data Sets in Science and Engineering: A Report to the National Science Foundation," (http://www.arl.org/bm~doc/digdatarpt.pdf).

3. Chair, DSpace program chair, 2nd International Conference on Open Repositories, 2007. http://www.openrepositories.org/2007/

4. NSDL activities: review panel participant, 2004; annual program proposal reviewer, 2006; and NSDL Sustainability Committee member, 2006 – present.

5. External review panelist, American Library Association Office of Accreditation, UCLA (2004) and San Jose State University (2007).

RECENT COLLABORATORS
Nelson Baker (Georgia Tech/DLPE); Delinda Buie (University of Louisville); Barbara Dewey (University of Tennessee); Toby Graham (University of Georgia); Martin Halbert (Emory); Ron Hutchins (Georgia Tech/OIT); John Leonard (Georgia Tech/CoE); Rick Luce (Emory); Gail MacMillan (Virginia Tech); Robert McDonald (UCSD Supercomputer Center); Merryll Penson (University System of Georgia); Greg Raschke (NCSU); Vicky Reich (Stanford); Katherine Skinner (Emory); Anthony Smith (University of Miami); Paula de Stefano (New York University); Richard Sweeney (New Jersey Institute of Technology); Charles Thomas (Florida Library Center for Library Automation); Aaron Trehub (Auburn University).

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

2005-	Founder and Co-Director, MIT CSAIL Decentralized Information Group
2005 -	Visiting Professor, University of Southampton, Electronics & Computer Science
2003 -	Principal Research Scientist, MIT Computer Science and Artificial Intelligence Laboratory
1998 -	Technology and Society Policy Director, World Wide Web Consortium
1994-8	Co-founder & Deputy Director, Center for Democracy and Technology
1991-4	Deputy Policy Director, Electronic Frontier Foundation

Closely Related Publications:

Weitzner, D. J., Beyond Secrecy: New Privacy Protection Strategies for Open Information Spaces. *IEEE Internet Computing* 11, 5 (Sep. 2007), 96-95.

Weitzner, Abelson, Berners-Lee, Feigenbaum, Hendler, Sussman, Information Accountability, MIT CSAIL Technical Report MIT-CSAIL-TR-2007 (13 June 2007)

Tim Berners-Lee, Wendy Hall, James Hendler, Nigel Shadbolt, Daniel J. Weitzner, Creating a Science of the Web. SCIENCE VOL 313, 11 AUGUST 2006.

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.

Other Publications:

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

Testimony before the US Senate Commerce Committee, Hearing on Online Privacy, 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)

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

Founding Board Member and Director, Web Science Research Initiative (WSRI). (2005 – present)

Member, National Academy of Sciences Study Committee, Technical and Privacy Dimensions of Information for Terrorism Prevention and Other National Goals (2006 - present)

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

Co-Chair, W3C Patents and Standards Interest Group (2005 - present)

Editorial Board, Foundations and Trends in Web Science (2004- present)

Editoral Board, Journal of Privacy Technology (2000 - present)

Chair, W3C Patent Policy Working Group (1999 – 2003)

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)

Selected Workshops

Co-Chair, Toward More Transparent Government: W3C/WSRI Workshop on eGovernment and the Web (2007)

Co-Chair, W3C Workshop on Semantic Web for Life Sciences (October 2004)

Co-Chair, W3C Wokshop on the Future of the Platform for Privacy Preferences (2001)

Selected Colloborators

H. Abelson (MIT), T. Berners-Lee (MIT), J. Hendler (RPI), J. Feigenbaum (Yale), L. Kagal (MIT), D. McGuinness (Stanford), E. Mogel (Columbia U.), N. Shadbolt (U. Southampton), G. Sussman (MIT)

Thesis Advisor and Postgraduate-Scholar Sponsor:

O. Seneviratne (PhD candidate, 2007-), H. Jones (M.Eng., 2006), Y. Scharf (M.Eng, 2007-), Lalana Kagal (Post Doctoral Fellow, 2005-07), Simson Garfinkel (PhD committee), Latanya Sweeney (PhD committee)

Selected Invited Lectures

Invited Essay, "Personal privacy without computational obscurity: Rethinking privacy protection strategies for open information networks," Annual Computer Security Applications Conference (13 December 20078)

Keynote, :"Challenges of the Policy Aware Web," Workshop on Privacy Enforcement and Accountability with Semantics.   6th International Semantic Web Conference (11 November 2007)

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 Febuary 2003)

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

Recent Teaching

MIT 6.805/6.806/STS085: Ethics and Law on the Electronic Frontier: Transparency, Privacy and Accountability, with H. Abelson (Fall 2007, 2005, 2004, 2003, 2002, 2000, 1999, 1998)

John Thompson Wilbanks

Science Commons / Creative Commons
Cambridge, MA 02139
[wilbanks@creativecommons.org]

BIOGRAPHICAL SKETCH

PROFESSIONAL PREPARATION

B.A., Philosophy, 1994, Tulane University

(includes a year of honors studies in Lettres Modernes at Universite de Paris-IV La Sorbonne)

APPOINTMENTS / EMPLOYMENT (current in **bold**)

2005-   Creative Commons, Vice President for Science Commons

2004-07 MIT, Computer Science and Artificial Intelligence Laboratory (Visiting Scientist)

2004   MIT, World Wide Web Consortium (Fellow, Semantic Web for Life Sciences)

2000-03 Incellico, Inc. (Founder and Chief Executive Officer)

1998-00 Harvard Law School, Berkman Center for Internet & Society (Assistant Director)

1997-98 fonix, inc (Research Analyst)

1996-97 United States House of Representatives, Office of Fortney "Pete" Stark (Legislative Assistant)

1994-96 American Physical Therapy Association, Government Affairs (Grassroots Coordinator)

Research Areas: Intersection of law and cyberspace, Semantic Web, science and technology policy, innovation systems, sharing economies, bioinformatics.

PUBLICATIONS

Most Closely Related Publications

2007   Wilbanks J. "Cyberinfrastructure for Knowledge Sharing" *CTWatch Quarterly Volume 3 Number 3* (August 2007)

2006   Wilbanks J. "Another reason for opening access to research" *British Medical Journal 333* (7582) , 1306 - 1308 (2006)

SYNERGISTIC ACTIVITIES

- Actively involved as Board member of relevant organizations (e.g. Board of Advisors for U.S. National Library of Medicine's PubMed Central, Board of Directors at Fedora Commons, Board of Directors at AcaWiki, Steering Committee for Open WetWare)
- Co-developed the Global Information Commons for Science Initiative (GICSI) with the Committee on Data and Technology for Science and Technology (CODATA), an interdisciplinary Scientific Committee of the International Council for Science (ICSU),. GICSI is a multi-stakeholder initiative arising from the second phase of the World Summit on the Information Society in Tunis in November 2005. Our other partners in the GICSI include the International Council for Scientific and Technical Information (ICSTI), the International Network for the Availability of Scientific Publications (INASP), World Data Centers (WDC), the International Council for Science (ICSU), the InterAcademy Panel on International Issues (IAP), the Academies of Science in Developing Countries (TWAS), the Organization for Economic Co-operation and Development (OECD), and the United Nations Economic, Scientific, and Cultural Organization (UNESCO).
- Founded and led the Neurocommons initiative. Brought together key stakeholders: private research funders in Huntington's Disease and other neurodegenerative disorders, prominent research scientists from US National Institutes of Health-funded projects, leaders from the Society for Neuroscience, the World Wide Web Consortium, Millennium Pharmaceuticals, the iBridge Network of technology transfer, and "open access" publishers BioMed Central, Hindawi, and PLoS One.
- International speaking authority on Open Access and Open Science (150+ invited presentations in 20 countries form 2005-present).

COLLABORATORS AND OTHER AFFILIATIONS

Collaborators

Hal Abelson (MIT), **Ben Adida** (Harvard), **Robi Blumenstein** (HighQ Foundation), **James Boyle** (Duke), **John Bracken** (MacArthur Foundation), **Matt Cockerill** (BioMed Central), **Paul David** (Oxford / Stanford), **Melina Fan** (Addgene), **Brian Fitzgerald** (Queensland University of Technology), **Melissa Hagemann** (Open Society Institute), **Heather Joseph** (Scholarly Publishing and Resources Coalition), **Eric Kansa** (California – Berkeley), **Tom Knight** (MIT), **Lesa Mitchell** (Kauffman Foundation), **Laura Paglione** (iBridge Network / Kauffman Foundation), **Paul Peters** (Hindawi Publishing), **Arti Rai** (Duke), **MacKenzie Smith** (MIT / DSpace), **Susie Stephens** (Eli Lilly), **Sir John Sulston** (Sanger Institute), **Ralph Swick** (World Wide Web Consortium), **Stacie Weninger** (Fidelity Foundation), **Gavin Yamey** (Public Library of Science)

	FOR NSF USE ONLY	
	PROPOSAL NO.	DURATION (months)
		Proposed | Granted
ORGANIZATION **Massachusetts Institute of Technology**	AWARD NO.	

A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)	NSF Funded Person-months			Funds Requested By proposer	Funds granted by NSF (if different)
	CAL	ACAD	SUMR		
1. **Stuart E Madnick - none**	0.00	0.00	0.00	$ 0	$
2. **Harold Abelson - none**	0.00	0.00	0.00	2	
3. **Jerrold M Grochow - none**	0.00	0.00	0.00	0	
4. **MacKenzie Smith - none**	0.00	0.00	0.00	0	
5.					
6. (**0**) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.00	0.00	0.00	0	
7. (**4**) TOTAL SENIOR PERSONNEL (1 - 6)	0.00	0.00	0.00	2	
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)					
1. (**0**) POST DOCTORAL SCHOLARS	0.00	0.00	0.00	0	
2. (**0**) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	0.00	0.00	0.00	0	
3. (**0**) GRADUATE STUDENTS				0	
4. (**0**) UNDERGRADUATE STUDENTS				0	
5. (**0**) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)				0	
6. (**0**) OTHER				0	
TOTAL SALARIES AND WAGES (A + B)				2	
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)				0	
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)				2	
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING $5,000.)					
TOTAL EQUIPMENT				0	
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)				0	
2. FOREIGN				0	

F. PARTICIPANT SUPPORT COSTS
1. STIPENDS $ 0
2. TRAVEL   0
3. SUBSISTENCE   0
4. OTHER   0

TOTAL NUMBER OF PARTICIPANTS (**0**) TOTAL PARTICIPANT COSTS	0	
G. OTHER DIRECT COSTS		
1. MATERIALS AND SUPPLIES	0	
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION	0	
3. CONSULTANT SERVICES	0	
4. COMPUTER SERVICES	0	
5. SUBAWARDS	0	
6. OTHER	0	
TOTAL OTHER DIRECT COSTS	0	
H. TOTAL DIRECT COSTS (A THROUGH G)	2	
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE) **(Rate: , Base:)**		
TOTAL INDIRECT COSTS (F&A)	0	
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)	2	
K. RESIDUAL FUNDS	0	
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)	$ 2	$
M. COST SHARING PROPOSED LEVEL $ **0** AGREED LEVEL IF DIFFERENT $		

PI/PD NAME **Stuart E Madnick**	FOR NSF USE ONLY		
	INDIRECT COST RATE VERIFICATION		
ORG. REP. NAME*	Date Checked	Date Of Rate Sheet	Initials - ORG

SUMMARY PROPOSAL BUDGET

Cumulative

	FOR NSF USE ONLY	
ORGANIZATION **Massachusetts Institute of Technology**	PROPOSAL NO.	DURATION (months)
		Proposed / Granted
PRINCIPAL INVESTIGATOR / PROJECT DIRECTOR **Stuart E Madnick**	AWARD NO.	

A. SENIOR PERSONNEL: PI/PD, Co-PI's, Faculty and Other Senior Associates (List each separately with title, A.7. show number in brackets)	NSF Funded Person-months			Funds Requested By proposer	Funds granted by NSF (if different)
	CAL	ACAD	SUMR		
1. **Stuart E Madnick - none**	0.00	0.00	0.00	$ 0	$
2. **Harold Abelson - none**	0.00	0.00	0.00	2	
3. **Jerrold M Grochow - none**	0.00	0.00	0.00	0	
4. **MacKenzie Smith - none**	0.00	0.00	0.00	0	
5.					
6. ( ) OTHERS (LIST INDIVIDUALLY ON BUDGET JUSTIFICATION PAGE)	0.00	0.00	0.00	0	
7. (**4**) TOTAL SENIOR PERSONNEL (1 - 6)	0.00	0.00	0.00	2	
B. OTHER PERSONNEL (SHOW NUMBERS IN BRACKETS)					
1. (**0**) POST DOCTORAL SCHOLARS	0.00	0.00	0.00	0	
2. (**0**) OTHER PROFESSIONALS (TECHNICIAN, PROGRAMMER, ETC.)	0.00	0.00	0.00	0	
3. (**0**) GRADUATE STUDENTS				0	
4. (**0**) UNDERGRADUATE STUDENTS				0	
5. (**0**) SECRETARIAL - CLERICAL (IF CHARGED DIRECTLY)				0	
6. (**0**) OTHER				0	
TOTAL SALARIES AND WAGES (A + B)				2	
C. FRINGE BENEFITS (IF CHARGED AS DIRECT COSTS)				0	
TOTAL SALARIES, WAGES AND FRINGE BENEFITS (A + B + C)				2	
D. EQUIPMENT (LIST ITEM AND DOLLAR AMOUNT FOR EACH ITEM EXCEEDING $5,000.)					
TOTAL EQUIPMENT				0	
E. TRAVEL 1. DOMESTIC (INCL. CANADA, MEXICO AND U.S. POSSESSIONS)				0	
2. FOREIGN				0	
F. PARTICIPANT SUPPORT COSTS					
1. STIPENDS $ 0					
2. TRAVEL 0					
3. SUBSISTENCE 0					
4. OTHER 0					
TOTAL NUMBER OF PARTICIPANTS (**0**) TOTAL PARTICIPANT COSTS				0	
G. OTHER DIRECT COSTS					
1. MATERIALS AND SUPPLIES				0	
2. PUBLICATION COSTS/DOCUMENTATION/DISSEMINATION				0	
3. CONSULTANT SERVICES				0	
4. COMPUTER SERVICES				0	
5. SUBAWARDS				0	
6. OTHER				0	
TOTAL OTHER DIRECT COSTS				0	
H. TOTAL DIRECT COSTS (A THROUGH G)				2	
I. INDIRECT COSTS (F&A)(SPECIFY RATE AND BASE)					
TOTAL INDIRECT COSTS (F&A)				0	
J. TOTAL DIRECT AND INDIRECT COSTS (H + I)				2	
K. RESIDUAL FUNDS				0	
L. AMOUNT OF THIS REQUEST (J) OR (J MINUS K)				$ 2	$
M. COST SHARING PROPOSED LEVEL $ **0** AGREED LEVEL IF DIFFERENT $					

PI/PD NAME **Stuart E Madnick**	FOR NSF USE ONLY
	INDIRECT COST RATE VERIFICATION
ORG. REP. NAME*	Date Checked / Date Of Rate Sheet / Initials - ORG

C *ELECTRONIC SIGNATURES REQUIRED FOR REVISED BUDGET

Budget Justification Page

This is a preliminary proposal that is to be submitted to NSF. If chosen for a full proposal submission, it has been proposed that some of the work will be conducted by the following subawardees: RENCI, HP Labs, Georgia Tech, Rice University and Harvard University. Appropriate subaward information will be provided in the official proposal due in March 21, 2008.

PI/Co-PI/ Senior Personnel	Current/ Pending	Agency	Title	Award Amount	Effective Date	End Date	Person Months per year	Location
Abelson, Hal	Current	NSF	CT-T: Transparent Accountable Datamining Intiative (TAMI)	1,370,000	2005-09-01	2008-08-31	0	MIT
Abelson, Hal	Current	NSF	Workshop on The Living Heritage of Artificial Intelligence	40,320	2007-08-01	2008-02-29	0	MIT
Madnick, Stuart	Current	Mubadala Development Company, PJSC	Semantics-Enabled Technology Forecasting: A Case Study on Alternative Energies	237,581	2006-12-01	2009-08-31	0	MIT
Madnick, Stuart	Current	Defense Science & Technology Agency	Context Knowledge Management Approach to Armament Safety Management	117,824	2007-04-01	2008-03-31	2.25	MIT
Madnick, Stuart	Current	Air Force	Identifying Unintended Consequences: Exploiting System Dynamics	690,425	2005-03-09	2008-03-31	0	MIT
Madnick, Stuart	Current	Air Force Research Laboratory	Proactive Intelligence (PAINT): Model Development	295,184	2007-07-24	2011-09-30	2.25	MIT
Madnick, Stuart	Current	Multi-Source	Productivity From Information Technology	70,543	1996-05-09	2008-06-30	0	MIT
Madnick, Stuart	Current	Multi-Sponsored Consortium	Total Data Quality Management (TDQM)	132,886	1992-09-01	2000-12-31	0	MIT
Madnick, Stuart	Current	Sequent Computer	Sequent Computer Systems	173,388	1994-04-24	2007-12-31	0	MIT
Madnick, Stuart	Current	National Security Innovations, Inc.	Sudan Stability Assessment Modeling (SSAM)	84,969	2007-04-20	2007-10-19	0	MIT
Madnick, Stuart	Pending	Lockheed Martin Advanced Technology Laboratories	Integrated Crisis Early Warning System (ICEWS)	560,000	2007-05-01	2008-04-30	0.9	MIT
Smith, MacKenzie	Current	Institute of Museum and Library Services	FACADE: Future-Proofing Architectural Computer-Aided Design	724,415	2006-10-01	2008-09-30	2.4	MIT
Smith, MacKenzie	Current	Andrew W. Mellon Foundation	Large-Scale data Interoperability Issues	1,530,136	2005-09-23	2008-09-30	1.8	MIT

<div align="center">

DataNet Preliminary Proposal: MIT DataSpace Project –
PI's and Potential Senior Personnel
& Brief description of each person's contribution to the MIT DataSpace Project

</div>

PI:

Stuart Madnick, John Norris Maguire Professor of MIT Information Technologies, MIT Sloan School of Management; Professor of Engineering Systems, MIT School of Engineering; co-Head, Total Data Quality Management (TDQM) Program; co-Head, MIT Productivity from Information Technology (PROFIT) Program: *Research on data integration, data re-use & re-purposing, data semantics and ontologies, data quality & provenance, data standards, data ownership/re-use legislation, very large data bases, sustainable business models.*

Co-PI's:

Hal Abelson, MIT Electrical Engineering & Computer Science Department & Computer Science and Artificial Intelligence Laboratory (CSAIL): *Technical, institutional and public policy questions necessary to advance the development of global, decentralized information environments.*

MacKenzie Smith, Associate Director for Technology, MIT Libraries: *Operations experience with open source data archiving software for broad access and long-term preservation (e.g. DSpace program) and research experience in federated, policy-based data management, Semantic Web models for data access and interoperability, and long-term preservation of new digital data*

Jerrold Grochow, VP Information Services & Technology, MIT: *Responsible for the management and operation of all information services at MIT and prior to joining MIT had 30 years of experience in technology management for government, industry and nonprofit organizations.*

Potential Senior Personnel*:

Timothy Berners-Lee, World Wide Web Consortium (W3C), Web Science Research Initiative (WSRI), and MIT Computer Science and Artificial Intelligence Laboratory (CSAIL): *Research on web architectural scalability,Semantic web research*

Pat Dreher, Renaissance Computing Institute (RENCI): *Storage architecture planning and prototyping*

John Erickson, HP Labs (Digital Media Systems Group): *Federation and policy-based curation, personalization*

Geneva Henry, Rice University**:** *Digital Library architecture and operation,library outreach for data curation*

Mei Hsu, HP Labs (Advanced Business Intelligence Lab): *Database and data warehouse research*

Henry D. Jacoby, MIT Center for Energy and Environmental Policy Research: *Energy / Environment domain expert*

David Karger, MIT Electrical Engineering and Computer Science Department & MIT Computer Science and Artificial Intelligence Laboratory (CSAIL): *Research on semantic Web technologies that improve access, management and reuse among digital assets, data visualization technologies*

Isaac Keohane, Harvard Medical School, Center for Biomedical Informatics**:** *expertise in computer science decision-support and research in machine-learning applied to biomedicine and translating genomic research into clinical practice*

Michele Kimpton, DSpace Foundation**:** *Large-scale archiving operations and federated digital repository management*

Tom Malone, MIT Sloan School of Management & MIT Center for Collective Intelligence (CCI): *Research on the use of scientific data and models to support collective decision making on a global scale*

Alexa McCray, Harvard Medical School, Center for Biomedical Informatics**:** *Expertise in digital libraries and the intersection of computer and information science and medicine.*

Jill Mesirov, The Broad Institute; Computational Biology, & Bioinformatics**:** *Life Sciences domain expert*

Ronald Prinn, MIT Department of Earth, Atmospheric and Planetary Sciences & MIT Center for Global Change Science: *Energy/Environment domain expert*

Michael Siegel, MIT Sloan School of Management & MIT Productivity from Information Technology (PROFIT) Program: *Research on data integration, data re-use & re-purposing, data semantics and ontologies, and MIT's Context INterchange (COIN) technologies.*

Michael Stonebraker, Vertica Corporation, StreamBase Systems, and, and MIT Computer Science and Artificial Intelligence Laboratory (CSAIL): *High-performance database system design and architectures, commercially viable technologies.*

Tyler Walters, Georgia Institute of Technology**:** *Digital Library architecture and operation,library outreach for data curation*

Danny Weitzner, World Wide Web Consortium (W3C), Web Science Research Initiative (WSRI) , and MIT Computer Science and Artificial Intelligence Laboratory (CSAIL): *Research on web architectural scalability, web policy models & community infrastructure to build global consensus on technical designs and community operating conventions*

John Wilbanks, Science Commons: *policy contracts and technology to increase access to research literature and materials, and increasing the utility of online data.*

***** *All of the Potential Senior Personnel listed have been contacted and have agreed to participate on this project. They are listed as "potential" because we were not sure that we could obtain the needed Financial Conflict of Interest materials for all of them in time due to the holidays and travel.*

Biographical sketches are not provided for the following

Proposed Initial DataSpace Advisory Board ******

Christine L. Borgman, Professor & Presidential Chair in Information Studies, Department of Information Studies, Graduate School of Education and Information Science, University of California, Los Angeles

Neil Geddes, Director**,** e-Science Centre, UK Science and Technology Facilities Council, Rutherford Appleton Laboratory

Keith Jefferies, President, European Research Consortium of Informatics and Mathematics (ERCIM): *Advice on common long-term scientific data storage needs and approaches from research organizations across the European Union*

Eric Lander, Director, Broad Institute; member, Whitehead Institute; Professor of Biology, MIT: *Life Sciences domain expert*

Neal Lane, Malcolm Gillis University Professor, James A. Baker III Institute for Public Policy, Rice University

Liz Lyon, UK Digital Curation Centre (DCC)

Ed Roberts, David Sarnoff Professor of Management of Technology, MIT Sloan School of Management; MIT Technological Innovation & Entrepreneurship Program; and MIT Entrepreneurship Center: *Advice on developing business model for long-term sustainability of DataNet*

Pam Samuelson, Professor, University of California at Berkeley School of Information and School of Law

Dan Schutzer, Director, Financial Services Technical Consortium (FSTC): *Advice on best practices from the financial services industry as well as advice on the potential for the financial services industry providing long-term support for DataNet-type systems and technology*

Raymie Stata, Chief Architect for Search and Marketplace, Yahoo.

Andrew Treloar, Director and Chief Architect, ARCHER Project, Australian National Data Service (ANDS), Monash University, Clayton, Australia.

*** Most of these people have been contacted and have agreed to serve on the Advisory Board.*