

Workshop on Information Integration Oct. 25 -27 2006, Philadelphia, USA

Attendees

David Archer Portland State University

I am currently a PhD student working in the area of information integration. My background includes 23 years in the computing industry: 7 years in custom CPU design; 8 years in system level design of workstations, servers, and supercomputers; and 8 years in engineering management and senior management roles. My particular area of interest is in capturing, making explicit, and exploiting the implicit information integration decisions we make as we use information for daily tasks.

Jason R. Baron National Archives and Records Administration (NARA)

Jason Baron has served for the past six years as Director of Litigation for the U.S. National Archives and Records Administration. Previously, Mr. Baron held the positions of trial attorney and senior counsel during twelve years with the Department of Justice in Washington, D.C., where he represented the interests of the U.S. government in a variety of complex lawsuits, including involving access to governmental information. Mr. Baron appeared as lead counsel in two leading cases filed against the White House and the Archivist of the United States involving electronic records management (*Armstrong v. Executive Office of the President* and *Public Citizen v. Carlin*). Among his publications, he is the author of a recent law review article entitled "Toward A Federal Benchmarking Standard For Evaluating Information Retrieval Products Used In E-Discovery," published in 6 *Sedona Conference Journal* 237 (2005). Mr. Baron serves as NARA's representative to the Sedona Conference Working Group on Electronic Records Retention and Production, where he co-chairs a special projects subgroup devoted to Search and Text Retrieval Science issues. He also currently serves as co-coordinator of the National Institute of Standards and Technology TREC 2006 Conference legal track. He has taught at the University of British Columbia, the University at Albany, and currently holds the position of Adjunct Professor at the University of Maryland's graduate College of Information Studies. He also previously participated as a member of the U.S. InterPARES team, an international research project studying issues related to electronic record preservation. Mr. Baron holds degrees from Wesleyan University and Boston University School of Law.

Philip Bernstein Microsoft Research

Philip A. Bernstein is a Senior Researcher at Microsoft Research and an Affiliate Professor at the University of Washington. In the distant past, he worked on database theory, query optimization and transaction processing and has co-authored two books on the latter topic. For the past decade, he has focused on meta data management -- mechanisms to manipulate schemas and mappings for

data integration. Dr. Bernstein is an ACM Fellow, a winner of the ACM SIGMOD Innovations Award, and a member of the National Academy of Engineering. He is on the boards of many journals, is Treasurer of the Computing Research Association, and is a member of the National Academy Board on Mathematical Sciences and Applications. He holds a B.S. from Cornell University and Ph.D. from University of Toronto. Additional details can be found at www.research.microsoft.com/~philbe.

Howard Bilofsky University of Pennsylvania

<http://www.pcbi.upenn.edu/cv/bilofsky.rtf>

Senior Fellow - School of Engineering and Applied Science, UPenn

Member of Scientific Advisory Board, HealthGrid Association

Long interest in Biomedical Informatics, Information Integration and Mediated and Grid Architectures.

Peter Buneman University of Edinburgh

Peter Buneman worked on databases for many years at the University of Pennsylvania. He now works at the University of Edinburgh, but takes every possible opportunity to visit his colleagues at Penn.

Michael Carey BEA Systems, Inc.

Michael J. Carey is currently the architect for the AquaLogic Data Services Platform product at BEA Systems, Inc. Mike manages the development team responsible for "all things XQuery" at BEA, including uses of XQuery for data integration (in ALDSP) and for application integration (in the data transformation engine of WebLogic Integration and the BEA AquaLogic Service Bus). Prior to joining BEA, his past lives include a stint at a Silicon Valley e-commerce startup, time at IBM Research working on DB2 and data integration related technologies, and a over a decade as a database faculty member in the Computer Sciences Department of the University of Wisconsin-Madison. He is an ACM Fellow and a member of the National Academy of Engineering.

**Robert Chadduck National Archives and Records Administration
(NARA)**

Mr. Robert Chadduck is the director of research for the Electronic Records Archives Program at the National Archives and Records Administration. Mr. Chadduck has worked in a variety of different domains and government programs. His background is in applied mathematics and oceanography (specifically, hydrodynamic modeling) and he also spent several years working for the Navy. At the National Archives his work focuses on information management, data architectures, and bridging different domains. Mr. Chadduck initiated a new research activity in long-term data preservation and access, and is the National Archives' first Networking and Information Technology Research and Development (NITR&D) appointee. He served as a reviewer for the National Academies Committee on Archiving and Accessing Environmental and Geospatial Data at NOAA's interim report, and has also consulted for the National Academies' Computer Science and Technology Board.

Surajit Chaudhuri Microsoft Research

Surajit Chaudhuri leads the Data Management and Exploration Group at Microsoft Research. In 1996, Surajit started the AutoAdmin project on self-tuning database systems at Microsoft Research and developed novel automated physical design tuning technology for SQL Server 7.0, SQL Server 2000 and SQL Server 2005. More recently, Surajit has begun work in the area of data cleaning and integration techniques. Part of this research has been incorporated in Microsoft SQL Server 2005. Surajit is also interested in the problem of querying and discovery of information in a flexible manner exploiting text search as well as DBMS querying functionality. Surajit did his Ph.D. from Stanford University in 1991 and worked at Hewlett-Packard Laboratories, Palo Alto from 1991-1995. He is an ACM Fellow and was awarded the 2004 ACM SIGMOD Contributions award for developing the Conference management Service at Microsoft Research.

Yi Chen Arizona State University

Yi Chen is an assistant professor in the Department of Computer Science and Engineering at Arizona State University. She received her Ph.D. and M.S. in Computer and Information Science from the University of Pennsylvania in 2005 and 2000, respectively, and her B.S. in Computer Science from Central South University, China in 1999. Her research interests include query evaluation and optimization techniques for database systems and data streams, data integration in the presence of uncertainty and incompleteness, data model and query language design, and scientific data management. For more information, please refer to her home page <http://www.public.asu.edu/~ychen127/>

Christopher G Chute Mayo Clinic College of Medicine

Dr. Chute received his undergraduate and medical training at Brown University, internal medicine residency at Dartmouth, and doctoral training in Epidemiology at Harvard. He is Board Certified in Internal Medicine, and a Fellow of the American College of Physicians, the American College of Epidemiology, and the American College of Medical Informatics. He became Head of the Section of Medical Information Resources at Mayo Foundation in 1988 and is now Professor and Chair of Biomedical Informatics. As a career scientist at Mayo, Dr. Chute's NIH and AHCPR/AHRQ funded research in medical concept representation, clinical information retrieval, and patient data repositories have been widely published. He is a Board member of the ANSI Health Information Standards Technology Panel, Convener of Healthcare Concept Representation WG3 within the ISO Health Informatics Technical Committee, chair-elect of the US delegation to ISO TC215 for Health Informatics, co-chair of the HL7 Vocabulary Committee, Chair of the Biomedical Computing and Health Informatics study section at NIH, and has served on or chaired multiple other NIH biomedical informatics study sections. He has chaired the International Medical Informatics Association WG6 on Medical Concept Representation since 1994 and was twice elected Vice-Chair of the ANSI Health Informatics Standards Board.

Sarah Cohen-Boulakia University of Pennsylvania

Sarah Cohen-Boulakia is a post-doctoral researcher at the University of Pennsylvania where she works with Prof. Susan Davidson. She defended her

PhD in Computer Science in 2005 at the Laboratoire de Recherche en Informatique, University of Paris-Sud 11, France. Dr. Cohen-Boulakia's research interests are in the design and application of integration systems dedicated to biological and biomedical domain. She is best known for her work on BioGuide and techniques for supporting biologists navigate the maze of biological resources available over the web.

In this work, she collaborates closely with biologists, physicians, and computer scientists. More information at <http://www.seas.upenn.edu/~sarahcb>.

Susan Davidson **University of Pennsylvania**

Dr. Davidson did her undergraduate work at Cornell University, and her PhD at Princeton University. She is the Weiss Professor of Computer and Information Science at the University of Pennsylvania, where she has been since 1982. Her research interests include database systems, data modeling, database integration, distributed systems, scientific databases and bioinformatics. Within bioinformatics she is best known for her work in data integration strategies.

AnHai Doan **University of Wisconsin-Madison**

AnHai Doan is an assistant professor in Computer Science at the University of Wisconsin, Madison. His interests cover databases and AI, with a current focus on data integration, information extraction, schema and ontology matching, and machine learning. Selected recent honors include the ACM Doctoral Dissertation Award in 2003, and CAREER Award in 2004. Selected recent professional activities include co-editing special issues on semantic integration for SIGMOD Records and AI Magazine, and co-chairing WebDB at SIGMOD-05 and the AI Nectar track at AAAI-06.

Tim Finin **University of Maryland, Baltimore County**

Tim Finin (<http://umbc.edu/~finin/>) is a Professor in the Department of Computer Science and Electrical Engineering at the University of Maryland, Baltimore County (UMBC). He has over 30 years of experience in the applications of Artificial Intelligence to problems in information systems, intelligent interfaces and robotics and is currently working on software agents, the semantic web, and mobile computing. He holds degrees from MIT and the University of Illinois and has also held positions at Unisys, the University of Pennsylvania, and the MIT AI Laboratory.

J. Michael Fitzmaurice **Agency for Health Research Quality (AHRQ)**

Before joining the Office of the Director, Dr. Fitzmaurice was Director, Center for Information Technology, Agency for Health Care Policy and Research until March 1998. He joined the Public Health Service in 1987 as Director of the National Center for Health Services Research and Health Care Technology Assessment, coming from the Health Care Financing Administration where he was Acting Director, Office of Research. Previously, as branch chief in the Office of Research, he directed the development of Medicare's Prospective Payment System. Dr. Fitzmaurice received his Ph.D. in Economics in 1972 from the University of Maryland. He has separate undergraduate degrees from St. Joseph's College (Rensselaer, IN) in Mathematics and in Economics. He was an adjunct Associate Professor in the Howard University Graduate School of

Business, 1972-77, and the University College Graduate School, University of Maryland at College Park, 1982-87.

He is active nationally and internationally in encouraging the uniformity and computerization of health care information to improve the quality of patient care and public health policy decisions. In 2000, he received the Elmer Gabrielli Award from the American Society for Testing and Materials, Committee E31, for his dedication to and achievements in national health data standards. In 1999, he received the Future of Health Technology Award, and was elected a Fellow of the American College of Medical Informatics.

In 1993, he served on the White House Health Reform Task Force (Information Systems and Administrative Simplification Working Groups). The White House Information Infrastructure Task Force released his paper "Health Care and the National Information Infrastructure (NII)" in Putting the Information Infrastructure to Work (May 1994) that presents the vision of how the NII can benefit the health care sector. From 1996 to 2000, he co-chaired the HHS Information and Cross-cutting Implementation Team that provided oversight for developing the health data standards requirements of the Health Insurance Portability and Accountability Act (HIPAA) of 1996 (P.L. 104-191). As a member of the HHS Privacy Working Group, he helped produce the HIPAA Privacy Rule. He was lead staff to the NCVHS Computerized Patient Record Working Group that produced the congressionally mandated report on standards needed for patient medical record information and its electronic transmission. This report was sent to the Secretary of HHS on July 6, 2000, and is available at <http://ncvhs.hhs.gov/hipaa000706.pdf>

Currently, Dr. Fitzmaurice advises the Director of AHRQ on developing health information technology research programs, and represents AHRQ on the White House Networking Information Technology Research and Demonstrations Program and on the National Committee on Vital and Health Statistics (NCVHS). From 2001-04, Dr. Fitzmaurice was lead staff on the HHS Secretary's Council on Private Sector Initiatives to Improve the Security, Safety, and Quality of Health Care. Dr. Fitzmaurice was also a member of the Office of the National Coordinator for Health Information Technology (ONCHIT) RFI Response Review Team that in 2005 reviewed and prepared summaries of the 500 public responses. <http://www.ahrq.gov/data/hipaa1.htm>. He initiated and now manages AHRQ's Data Standards Program, with an annual budget of \$10 million in 2004, 2005, and 2006.

**James C. French Science Applications International Corporation
(SAIC)**

From 2002 through 2006 I was on a leave of absence from the University of Virginia serving as a program director in the Division of Information and Intelligent Systems, Directorate for Computer and Information Sciences and Engineering at the National Science Foundation. I helped start and co-directed the Science and Engineering Information Integration and Informatics (SEIII) program. I am presently on a leave of absence from U.Va. working for Science Applications International Corporation where I am part of a group looking at innovative applications of biometrics and biotechnology. Information integration

and data fusion are fundamental elements of this kind of work.

Lise Getoor

University of Maryland

Assistant Professor, Computer Science Department, University of Maryland, College Park

Robert Grossman

University of Illinois at Chicago

Robert Grossman is the Director of the National Center for Data Mining (NCDM) and the Laboratory for Advanced Computing at the University of Illinois at Chicago. The NCDM performs research; manages the Teraflow Testbed, an international data mining testbed; operates the Data Mining Group, which develops the Predictive Model Markup Language (PMML) and related standards; and engages in outreach activities in the areas of data mining and data intensive computing.

Laura Haas

IBM Almaden Research Center

Laura Haas is an IBM Distinguished Engineer and Director of Computer Science at Almaden Research Center. Most recently, she was responsible for Information Integration Solutions (IIS) architecture in IBM's Software Group, after leading the IIS development team through its first two years. Dr. Haas joined the development team in 2001 as manager of DB2 UDB Query Compiler development. Previously, Dr. Haas was a research staff member and manager at IBM's Almaden Research Center for nearly twenty years. In Research, she worked on and managed a number of exploratory projects in distributed database systems. She is best known for her work on the Starburst query processor (from which DB2 UDB was developed), on Garlic, a system which allowed federation of heterogeneous data sources, and on Clio, the first semi-automatic tool for heterogeneous schema mapping. Garlic technology married with DB2 UDB query processing is the basis for WebSphere Information Integrator's federation capabilities, while Clio capabilities are a core differentiator for the new Rational Data Architect. Dr. Haas is an active member of the database community, serving as vice chair of ACM SIGMOD from 1989-1997, and, currently, as Vice President of the VLDB Board of Trustees, as well as on many program committees for technical conferences. She has received several IBM awards for Outstanding Technical Achievement, and an IBM Corporate Award for her work on federated database technology. She is a member of the IBM Academy of Technology.

Martin Herman

**National Institute of Standards and Technology
(NIST)**

Dr. Martin Herman is Chief of the Information Access Division, National Institute of Standards and Technology (NIST). He is responsible for the overall program in research, measurements, testing, and standards in information access technologies at NIST, including speech processing and human language technologies, multimedia information access, information retrieval, image and video processing technologies, biometrics, visualization and usability testing, human-computer and human-robot interfaces, smart spaces, and optical media testing. He has worked with the intelligence community, the homeland security and law enforcement communities, industry, and the research community on these topics.

Howard Ho**IBM Almaden Research Center**

Howard (Ching-Tien) Ho received a B.S. degree in electrical engineering from the National Taiwan University in 1979 and M.S., M.Phil., and Ph.D. degrees in computer science from Yale University in 1985, 1986, and 1990, respectively.

He joined IBM Almaden Research Center as a Research Staff Member in 1989. He was manager of the Foundations of Massively Parallel Computing group from 1994 to 1996, where he led the development of collective communication, as part of IBM MPL and MPI, for IBM SP-1 and SP-2 parallel systems. He is manager of the Intelligent Information Integration group and the Clio Project since March 2001. His current research interests include XML and database, especially on the schema mapping and information integration, and data mining. His past research interests include on-line analytical processing (OLAP), communication issues for interconnection networks, algorithms for collective communications, graph embeddings, fault tolerance, and parallel algorithms and architectures. He has published over 30 journal papers and over 50 conference papers in these areas.

Dr. Ho is a co-recipient of the 1986 "Outstanding Paper Award" of the International Conference on Parallel Processing. He has received an IBM Outstanding Innovation Award, two IBM Outstanding Technical Achievement Awards, and five IBM Plateau Invention Achievement Awards. He has filed 15 patents. He has served on the Editorial Boards of the IEEE Transactions on Parallel and Distributed Systems for four years. He was an Area Editor of the Journal of Interconnection Networks for three years and has been one of three Managing Editors for three years. He has served as program vice-chairs twice for parallel processing conferences. He has co-edited a book "Large-Scale Parallel Data Mining" by Springer-Verlag in 2000. He has served Program Co-Chair of the Workshop on the Large-Scale Parallel KDD Systems, 1999. He has served on over 30 program committees of conferences and workshops in parallel processing, data mining and database. He is a member of the ACM and the IEEE Computer Society.

Eduard Hovy**Information Sciences Institute, University of Southern California (ISI/USC)**

Eduard Hovy leads the Natural Language Research Group at the Information Sciences Institute of the University of Southern California. He is also Deputy Director of the Intelligent Systems Division, as well as a research associate professor of the Computer Science Department of USC and Advisory Professor of the Beijing University of Posts and Telecommunications. He completed a Ph.D. in Computer Science (Artificial Intelligence) at Yale University in 1987. His research focuses on information extraction, automated text summarization, question answering, the semi-automated construction of large lexicons and ontologies, machine translation, and digital government. Dr Hovy regularly serves in an advisory capacity to funders of NLP research in the US and EU. He is the author or co-editor of five books and over 170 technical articles. In 2001 Dr. Hovy served as President of the Association for Computational Linguistics (ACL) and in 2001–03 as President of the International Association of Machine Translation (IAMT). Dr. Hovy regularly co-teaches a course in the Master's

Degree Program in Computational Linguistics at the University of Southern California, as well as occasional short courses on MT and other topics at universities and conferences. He has served on the Ph.D. and M.S. committees for students from USC, Carnegie Mellon University, the Universities of Toronto, Karlsruhe, Pennsylvania, Stockholm, Waterloo, Nijmegen, Pretoria, and Ho Chi Minh City.

URLs:

<http://www.isi.edu/natural-language/nlp-at-isi.html>

<http://www.isi.edu/~hovy.html>

Sally Howe

National Coordination Office (NCO)

Sally E. Howe is Associate Director of the National Coordination Office (NCO) for Information Technology Research and Development (IT R&D), and has been at the NCO since it was established in 1992, serving first as Assistant Director for Technology and then as Chief of Staff.

Todd Hughes

Defense Advanced Research Projects Agency (DARPA)

Dr. Todd Hughes became a Program Manager for DARPA/IXO in July 2006. His research interests include geospatial intelligence, cognitive systems, and data integration. Prior to joining DARPA, Dr. Hughes held a position as Technical Manager at Lockheed Martin Advanced Technology Laboratories, where he oversaw research and development projects in cognitive systems and semantic technology—many of which were supported by DARPA. For example, he was the Project Manager for the Polymorphous Cognitive Agent Architecture project on the DARPA Architectures for Cognitive Information Processing program. In addition, he served as Principal Investigator for the Dynamic Trust-based Resources project on the DARPA Fault Tolerant Networks program. Other notable research endeavors include the Ontology Translation Protocol, a project aimed at developing technology to assist in the rapid integration of heterogeneous information resources. In the context of this initiative, Dr. Hughes collaborated with the National Institute of Standards and Technology to organize the first public evaluation of data model alignment tools: the Information Interpretation and Integration Conference. Dr. Hughes earned his Ph.D. from the University of Wisconsin-Madison in 2001. While completing his doctorate, he worked as a member of the technical staff at Cycorp, Inc.

Yannis Ioannidis

University of Athens, Department of Informatics

Yannis Ioannidis is currently a Professor at the Department of Informatics and Telecommunications of the University of Athens.

He received his Diploma in Electrical Engineering from the National Technical University of Athens in 1982, his MSc in Applied Mathematics from Harvard University, and his Ph.D.

degree in Computer Science from the University of California at Berkeley in 1986. Immediately after that he joined the faculty of the Computer Sciences Department of the University of Wisconsin at Madison, where he became a Professor before finally leaving in 1999. His research interests include database and information systems, digital libraries, personalization, scientific systems and workflows, eHealth systems, and human-computer interaction, topics on which

Science from Purdue University in 1991 and 1993 respectively. His research interests are in the broad area of networked computing and intelligent systems. His primary focus has been on data management and security for mobile, pervasive, and sensor systems. He has created agent based middleware to support discovery, composition, and secure access of services/data over both infrastructure based and ad-hoc wireless networks, as well as systems that integrate sensors with the grid. He is also interested in Semantic Web and Data/Web Mining, where he has worked on creating personalized and secure web spaces using a combination of agents, policies, and soft computing.

He has published over 125 technical papers, and has obtained research support from NSF, NASA, DARPA, DoD, IBM, AetherSystems, HP, AT&T and Intel. He has presented tutorials in conferences, served as guest editor for special issues for IEEE Personal Comm., Comm. ACM etc., and served as an Associate Editor of IEEE Transactions of Fuzzy Systems from 99-03. He currently serves on the editorial board of the Intl. J. Semantic Web and Information Systems. At UMBC, Joshi teaches courses in Operating Systems, Mobile Computing, Networking, Security, and Web Mining. He is a member of IEEE, IEEE-CS, and ACM.

Subbarao Kambhampati Arizona State University

<http://rakaposhi.eas.asu.edu/rao.html>

Junhyong Kim Dept. Biology, University of Pennsylvania

Junhyong Kim started out as a Microbiology major at Seoul National University and received his BS in 1984 on the topic of numerical taxonomy of marine flora. He moved to State University of New York at Stony Brook for graduate work in the Department of Ecology and Evolution, where he worked with Dr. Robert Sokal and F. James Rohlf on computational phylogenetics, with Dr. Dan Dykhuizen on E. coli competition, and with Dr. Lev Ginzburg on theoretical models of spatial population dynamics. He received his PhD in 1992 on the topic of influence of spatial structure on inhibitory competition dynamics. He continued on to postdoctoral training on Drosophila chromosome evolution with Dr. Margaret Kidwell at the University of Arizona. In 1994 he started as an Assistant Professor in the Department of Biology at Yale University working on the evolution of gene expression timing in Drosophila early development. He also continued to work on computational biology problems including computational phylogenetics, multi-dimensional gene expression analysis, and genomic annotation. In 1999, with John Carlson at Yale, he helped discover olfactory receptor proteins in insects using computational annotation of the Drosophila genome. He received tenure in 2001. In 2002, he moved to University of Pennsylvania as Edmund J. and Louise W. Kahn Term Endowed Professor in Biology. Currently, he is an executive member of Penn Center for Bioinformatics and the co-Director of Penn Genomics Institute. His current work includes computational phylogenetics, computational and functional neurogenomics, and comparative genomics of yeast cell cycles. On the administration side, he founded the Computational Biology and Bioinformatics Graduate Track at Yale University, is an Associate Editor of IEEE/Transactions in Computational Biology and Bioinformatics, is on the Steering Committee of Workshop in Algorithms for

Bioinformatics (WABI), is on the Scientific Advisory Board of NESCENT, is a member of NRC committee on Conceptual Basis of Biology, and executive committee member of CIPRES.

Frankie Denise King National Coordination Office for Networking and Information Technology Research and Development (NCO/NITRD)

Frankie Denise King is the Technical Coordinator for the Human Computer Interaction and Information Management (HCI&IM) Coordinating Group (CG), one of seven formal interagency coordinating groups of the Networking and Information Research and Development Subcommittee(NITRD) of the National Science and Technology Council (NSTC), Executive Office of the President.

Craig Knoblock University of Southern California (USC)

Dr. Craig Knoblock is a Senior Project Leader at the Information Sciences Institute and a Research Professor in Computer Science at the University of Southern California (USC). He is also the Chief Scientist for both Fetch Technologies and Geosemble Technologies, which are spinoff companies from USC. He received his Ph.D. in Computer Science from Carnegie Mellon in 1991. His current research interests include information integration, automated planning, machine learning, and constraint reasoning and the application of these techniques to geospatial data integration.

He has published one book and over 150 articles, book chapters, and conference papers in information integration, planning, and machine learning. He served on the Senior Program Committees of the 1997, 1998, 2000, 2004, and 2006 National Artificial Intelligence Conferences, the Senior Program Committee of the 2006 International Joint Conference on AI, and the Senior Program Committee of the 2004 International Semantic Web Conference. He is currently the President of the International Conference on Automated Planning and Scheduling and a Fellow of the American Association of Artificial Intelligence.

Hilmar Lapp NESCent

Education:

Diploma (M.S.) in Biology (major) and Computer Science (minor), University of Freiburg, 1997

Employment/Experience:

1997 - 2001 Staff Scientist, Laboratory Head of Computational Biology & Data Mining, Novartis Research Institute, Vienna, Austria

2000 - 2001 Sabbatical, Genomics Institute of the Novartis Research Foundation (GNF), San Diego, CA

2001 - 2006 Staff Scientist, Group Leader Bioinformatics, Genomics Institute of the Novartis Research Foundation (GNF), San Diego, CA

2006 - present Assistant Director f. Informatics, National Evolutionary Synthesis Center (NESCent), Durham, NC

Other activities:

1999 - present Core developer of Bioperl, Lead developer of Bioperl-db and BioSQL

2001 - present Board of Directors, Open Bioinformatics Foundation (OBF)

Project with Phil Bernstein.

Renée J. Miller **University of Toronto**
Data Integration.

Partha Mitra **Cold Spring Harbor Laboratory**
Following graduate work in theoretical physics (Harvard, 89-93) Partha has worked in quantitative neuroscience and theoretical engineering, at Bell Laboratories ('93-'03). He works collaboratively with groups at Caltech ('96), NYU, Princeton and Cornell University Medical School. In the fall of 2003 Partha moved his research effort to Cold Spring Harbor Laboratory, where he is currently based.

Douglas Oard **University of Maryland**
Douglas W. Oard is Associate Dean for Research at the College of Information Studies, University of Maryland, College Park, where he holds joint appointments as an Associate Professor in the College of Information Studies and the Institute for Advanced Computer Studies, with affiliate appointments in the Computer Science Department and the Applied Mathematics and Scientific Computing Program. He holds a Ph.D. in Electrical Engineering from the University of Maryland, and his research interests center around the use of emerging technologies to support information seeking by end users. His recent work has focused on cross-language information retrieval, searching spoken and written conversational media, legal "e-discovery," and leveraging observable behavior to model interest and intent.
Additional information is available at <http://www.glue.umd.edu/~oard/>.

Yannis Papakonstantinou **University of California, San Diego**
Yannis Papakonstantinou is an Associate Professor of Computer Science and Engineering at the University of California, San Diego. His research is in the intersection of database and Internet technologies. Yannis has published over sixty research articles in scientific conferences and journals, given tutorials at major conferences, and served on journal editorial boards and program committees for numerous international conferences and symposiums. He was the co-Chair of the ACM-sponsored WebDB 2002, the General Chair of ACM SIGMOD 2003, the Vice PC Chair for the "XML, Metadata and Semistructured Data" track of IEEE ICDE 2004 and IEEE ICDE 2006, and the founding co-chair of the ACM-sponsored 1st Workshop on XQuery Implementation, Experience and Perspectives 2004. In 1998, Yannis received the NSF CAREER award. In 2000 Yannis founded Enosys Software, which built one of the first two generally available distributed XQuery processors, along with software for XML-based Enterprise Information Integration. Enosys Software was acquired in 2003 by BEA Systems, while the Enosys product was available under the BEA Liquid Data (now Aqualogic) brand name since 2002. Yannis holds a Diploma of Electrical Engineering from the National Technical University of Athens and MS and Ph.D. in Computer Science from Stanford University (1997).

Louiqa Raschid **University of Maryland**
Louiqa Raschid is a Professor at the University of Maryland in College Park. Her

research interests are focused around issues of information (data and knowledge) management, semantics and data integration, and performance. She investigate applications in the life sciences, health information systems, humanitarian IT applications and Grid computing. Her research utilizes a variety of methodologies and reference sub-areas such as optimization, large scale simulation using trace data, semantics and logic based reasoning, and data analysis techniques.

Arnon Rosenthal The MITRE Corporation

Arnon Rosenthal works at the MITRE Corporation, Bedford, MA, USA. His work in recent years has focused on administration problems for data interoperability and data security. Our group often attempts to help database researchers understand the pragmatic needs of administrators and developers. His earlier interests included query processing, database design, active databases, and discrete algorithms.

Len Seligman The MITRE Corporation

Len Seligman is a Senior Principal Scientist in MITRE's Center for Innovative Computing and Informatics. He has over 25 years experience researching, architecting, implementing, and managing advanced information systems. Since 1997, he has been Associate Editor (Industry Perspectives) of SIGMOD Record, the bulletin of the ACM Special Interest Group on the Management of Data. Dr. Seligman holds a Ph.D. in Information Technology from George Mason University and a Master's in Computer Science from the University of Illinois. His current research focus is information interoperability in large, dynamic enterprises. Dr. Seligman provides strategic guidance to diverse customers on information management challenges.

Home page: <http://www.mitre.org/staffpages/seligman/>

Jeffrey L. Solka Naval Surface Warfare Center Dahlgren Division (NSWCDD)

Jeffrey L. Solka was born in Harrisonburg, Virginia, on January 31, 1955. He earned the B.S. degree in Mathematics and Chemistry from James Madison University in 1978, the M.S. in Mathematics from James Madison University in 1981, the M.S. in Physics from Virginia Polytechnic Institute and State University in 1989 and his Ph.D. in Computational Sciences and Informatics (Computational Statistics) at George Mason University, working under the direction of Prof. Edward J. Wegman, in May of 1995. Since 1984, Dr. Solka has been working in nonparametric estimation and statistical pattern recognition for the Naval Surface Warfare Center, Dahlgren, VA. He has published over 120 journal, conference, and technical papers, has won numerous awards, and holds 4 patents.

Sylvia J Spengler National Science Foundation (NSF)
Program Director, Information Integration and Informatics

Dan Suci University of Washington

Dan Suci is an associate professor in Computer Science at the University of Washington. Professor Suci received his Ph.D. from the University of Pennsylvania in 1995, then was principal member of the technical staff at AT&T

Labs until he joined the University of Washington in 2000. Professor Suci is conducting research in data management, with an emphasis on topics that arise from sharing data on the Internet, such as management of semistructured and heterogeneous data, data security, and managing imprecisions in data. He is a co-author of the book *Data on the Web: from Relations to Semistructured Data and XML*, holds six US patents, received the 2000 ACM SIGMOD Best Paper Award, is a recipient of the NSF Career Award and of an Alfred P. Sloan Fellowship. Professor Suci conducted in the past three projects that lead to popular software tools in the public domain: XMill, XMLTK, and SilkRoute.

Simon Szykman **National Coordination Office for Networking and Information Technology Research and Development (NCO/NITRD)**

Dr Szykman is the Director of the National Coordination Office for Networking Information Technology Research and Development.

Wang-Chiew Tan **University of California, Santa Cruz**

Wang-Chiew Tan is an assistant professor at UC Santa Cruz. Her research interests include data provenance and information integration.

Val Tannen **University of Pennsylvania**

Research Interests include database programming languages, query optimization, distributed query processing, integration of heterogenous data sources, finite model theory, formal specification, data parallelism, integration of molecular biology and phylogenetic data sources.

Kenneth Thibodeau **National Archives and Records Administration (NARA)**

Kenneth Thibodeau is Director of the Electronic Records Archives (ERA) Program at the National Archives and Records Administration (NARA). An internationally recognized expert in electronic records, he has served as Chief of the Records Management Branch of the National Institutes of Health, and Director of the DoD Records Management Task Force, which developed the U.S. Government's standard for records management software. He earned a Ph.D. in the history and sociology of science from the University of Pennsylvania, and held post-doctoral fellowships in computer science at the University of Kansas and the University of Illinois. A Fellow of the Society of American Archivists, he has published over 30 papers and spoken at more than 150 conferences around the world.

Gary Walter **U.S. Environmental Protection Agency (EPA)**

Gary Walter, Computer Scientist, Program Manager, EPA Office of Research and Development. Interests include large scale data management, integration and fusion, and visualization techniques for high end computing environments.