

Alan Karr Named Director

Alan F. Karr has been appointed Director of NISS, effective July 1, 2000.

Announcing the appointment, Jon Kettenring, Chair of the Board of Trustees, said "Karr's selection, the culmination of a nationwide search, signals our readiness to take NISS to new levels of accomplishment and involvement with the statistics community."

Accepting, Karr stated that "I am honored to become Director of NISS, and extremely grateful to the Board of Trustees for their confidence in my ability to lead NISS. We have come a long way in nine years, but now we face new challenges to ensure that NISS becomes an Institute that the entire statistical sciences community is proud of. My goal is to work with the Board of Trustees and the community to meet these challenges."

Karr succeeds Jerome Sacks, founding Director of NISS. Through December 31, 2000, Sacks will assist with management of the NISS affiliates program, as well as remain involved in NISS research on transportation and computer model evaluation. Effective January 1, 2001, he will become Professor of Statistics and Decision Sciences at Duke University on a full-time basis. "On behalf of the Board of Trustees," Kettenring said, "I thank Jerry for having had the vision to start NISS and the wisdom to guide its development from a concept into a strong national resource in the mathematical sciences."

As Associate Director of NISS since 1992, Karr has developed and led research projects in transportation, software engineering (see page 4) and digital government. Prior to coming to NISS, he was Professor of Mathematical Sciences and Associate Dean of the School of Engineering at Johns Hopkins.

Affiliates Program Established Kickoff Meeting Held

A newly created program provides corporate, government and university affiliates opportunities to guide and engage in NISS research, and through NISS, to work with one another on shared interests. The program is off to an energetic start: to date, more than twenty corporate/government affiliates and nearly fifteen university affiliates have joined. A kickoff meeting was held at NISS on March 3, 2000. Plans are underway for rapid response on several problems of broad importance to affiliates, whose scale and complexity demand cross-disciplinary, cross-institutional efforts characteristic of NISS.

As Jon Kettenring, Chair of the NISS Board of Trustees, said at the kickoff meeting, "The affiliates program represents a strong new dimension for NISS. It will help define future directions and activities that are of value to affiliates as well as consistent with the overall mission of NISS." Jerome Sacks, Director of NISS, noted that in addition to supporting NISS efforts to foster cutting edge research on topics of direct interest to affiliates, the program will "provide access to NISS personnel, including postdoctoral fellows and graduate students, and will ensure strong contact between NISS researchers and those who own the problems and data."

Mary Ellen Bock (Purdue), who along with Kettenring, Sacks and George Williams (Merck) played a lead role in establishing the program, told university affiliates at the meeting that "this is a unique opportunity for faculty, postdocs and graduate students, through NISS, to work on important problems in government and industry." The program also allows statistical sciences departments and programs throughout the United States to learn rapidly of emerging

Continued on page 5

Director's Corner

Alan F. Karr



Over the next several months, in a variety of settings, I will articulate to the statistical sciences community a vision for NISS, and begin working with the community to refine the vision and move NISS forward. Here is the high-level vision: NISS exists, I believe, to enlarge the future of statistics. To serve the community in this way, NISS must be energized, intense and exciting. We must aim to make it one of the very best statistics research organizations in the country -- not best of its size or best of its kind but simply the best.

Three "grand challenges" arise in achieving the high-level vision. First, NISS research programs must be of the highest quality, with long-term depth in selected areas, rapid response exploration in others, strong attention to the professional development of the people involved and multiple paths for engagement. Even if not all who want to participate are able to, everyone must know that the opportunity is there.

Second, financial stability must be generated. A financially precarious NISS cannot serve the community effectively.

Finally, and my highest initial priority as Director, is a positive image for NISS. NISS must be, and be seen as, supportive of the statistical sciences community. Our mission is to "identify, catalyze and foster high-impact cross-disciplinary research involving the statistical sciences." Our impact, therefore, is not only what we do, but also how those benefit who participate in our programs (as senior researchers, postdocs, graduate and undergraduate students) or in activities that we catalyze. We should, and will, be proud of activities that we stimulate, but in which NISS plays no direct role.

This vision is attainable, but not by me alone. I invite you to join me in defining and realizing a strong future for NISS. Your ideas and opinions are valuable to me, and will be listened to. Please contact me at karr@niss.org or 919.685.9300.

NISS Postdocs

Jaeyong Lee will assume a faculty position in the Department of Statistics at Pennsylvania State University in August. For the past year, he has been involved in research on digital government, focusing on statistical consequences of geographical aggregation procedures invoked to preserve data confidentiality.

Li Liu will arrive at NISS at October, 2000. Currently, she is completing a Ph.D. (Statistics) at the University of Illinois at Urbana-Champaign. She anticipates being involved in research on computer model evaluation.

Byungkyu Park (Ph.D. Transportation Engineering, Texas A&M) has been at NISS since 1999. His research, under the transportation project, addresses use of traffic simulations to evaluate and design signalization plans for urban street networks, and associated issues of model evaluation.

Jennifer Pittman arrived in May, 2000. Jennifer received her Ph. D. (Statistics) from Pennsylvania State University. At NISS, she participates in collaborative research with statisticians and scientists from GlaxoWellcome on the "three-way problem," as well as in research on computer model evaluation.

Ashish Sanil (Ph.D. Statistics, Carnegie Mellon), who has been at NISS since 1998 and will remain for another year, is working on the digital government project, developing aggregation methods to disclose information derived from Federal databases, but maintain confidentiality of the data themselves.

NISS Newsletter

Published quarterly by the

National Institute of Statistical Sciences
PO Box 14006
Research Triangle Park, NC 27709
919.685.9300
admin@niss.org

© 2000, National Institute of Statistical Sciences
All Rights Reserved

NISS Calendar

Monday - Tuesday, July 10-11, 2000

NCES Workshop on NAEP Inclusion Strategies (Research Triangle Park)

Thursday, July 13, 2000

Affiliates Workshop on Gene Expression Data (Research Triangle Park)

Sunday, August 13, 2000, 11:00 AM - 2:00 PM

Affiliates Meeting (Cameral Room, Westin Hotel, Indianapolis)

Monday, August 14, 2000, 5:30 - 7:00 PM

NISS JSM 2000 Reception (Sangamore 6, Convention Center, Indianapolis)

Wednesday, August 16, 2000, 2:00 - 3:50 PM

NISS JSM 2000 Session on *Statistics and Information Technology* (Indianapolis)

Matthias Schonlau: Models for ISP Customer Churn

Ashish Sanil: Using the Web to Disseminate Information

but Preserve Confidentiality of Data

Stephen G. Eick and Alan F. Karr: Visual Scalability

Friday - Saturday, November 3-4, 2000

Board of Trustees Meeting (Research Triangle Park)

Thursday - Friday, November 30 - December 1, 2000

Affiliates Workshop on Data Quality (Morristown, NJ)

NISS-Related Talks at JSM 2000

Monday, August 14, 2000, 2:00 PM

Latanya Sweeney: Computer Science Tools, Confidentiality and Privacy

Wednesday, August 16, 2000, 12:30 PM

Todd Graves: Publishing Data Analyses on the Web

Wednesday, August 16, 2000, 2:45 PM

Jaeyong Lee: On Posterior Consistency of Survival Models

Thursday, August 17, 2000, 10:55 AM

Alan F. Karr: The NISS Digital Government Project

Research Highlights

NISS Builds TRANSIMS Activity Generator

The Los Alamos National Laboratory (LANL) is nearing completion of TRANSIMS, a simulation model that represents regional-level transportation from people who live there to the pollution consequences of the traffic they generate. NISS is responsible for the *activity generation* module, which synthesizes, for each person in the region, the daily activities that generate the need to travel.

Paul Speckman of the University of Missouri led the effort, together with the late Eric Pas (Duke), Dongchu Sun (Missouri) and Kenneth Vaughn (ex-NISS postdoc, now at the Metropolitan Transportation Commission in Oakland, CA).

The testbed for development of the activity generator was Portland, OR. Beginning with a population with detailed household characteristics (synthesized from Census data) and a survey of two days' of activities from 4,000 households in Portland, the activity generator first creates a pattern of activities (type, sequence, desired duration) for each member of each synthetic household, by resampling from demographically similar households in the survey. Then, locations for activities are generated using Census data on locations of schools, places of employment and shopping.

Statistical challenges were the scale of the activities to be generated (leading to millions of trips per day), capturing dependences among patterns for different individuals (for example, when there are shared rides) and accommodating feedback from other TRANSIMS modules that generate routes by individual travellers and the traffic and emissions consequences.



President Clinton is briefed about TRANSIMS during a visit to LANL.

Code Decay Project Concludes Successfully

The NISS software engineering project *Code Decay in Legacy Software Systems* reached its completion on April 30, 2000. This Lucent Technologies and NSF-funded collaboration of software engineers, statisticians and organization theorists produced important insights into the evolution and maintenance of large software systems.

Alan Karr (NISS) and Stephen Eick (Lucent) directed the project. Other key participants were Todd Graves (ex-NISS postdoc, now at Los Alamos), J. Stephen Marron (UNC Chapel Hill), Audris Mockus (Lucent), Adam Porter (Maryland), Nancy Staudenmayer (Duke), Harvey Siy (Lucent), Lawrence Votta (Lucent) and David Weiss (Lucent).

The goals of the project were to quantify, measure, predict and reverse or retard *code decay* -- the increasing difficulty and cost to change software over time. The project team treated code decay as a scientific, data-driven subject, using data representing the entire change history for Lucent's 5ESS telephone switch.

Selected highlights of the research:

Code decay indices that quantify and predict decay. The indices show, for example, that earlier changes induce decay, while sizes of changes, code complexity and activity of multiple developers (all suspects for decay) do not.

Statistical models for evolution of modularity of code, which show that modularity (a characteristic of good software) does break down over time, and were also applied to designate portions of the code for development at international sites.

Visualization methods for software changes, able to deal with the unusual scale (gigabytes) and nature (primarily text) of software change data.

Live Documents (US Patent 5, 937,064), tools that create new ways for readers of scientific documents to interact with the underlying data, for example, by performing alternative statistical analyses not presented by the authors of the document.

Continued from page 1

needs for statistics education.

Current affiliates are listed on page 6. Additional affiliates are actively being sought. For information, consult the affiliates home page on the NISS Web site -- www.niss.org/affiliates/affiliatesmain.html -- or contact Jerome Sacks at NISS (sacks@niss.org).

The kickoff meeting was attended by representatives of thirty charter affiliates. It featured eight presentations by corporate and government affiliates on potential areas of activity by the program. A report on the meeting, as well as many of the presentations, is available on the affiliates home page.

As a result of the meeting and subsequent discussions between NISS and affiliates, two topic areas have been identified for initial activity. The first, Data Quality, will be pursued by means of a workshop (organized jointly by NISS and Telcordia) on November 30 - December 1, 2000, whose goal will be to give form (and therefore a path for action) to this important problem, especially for large, complex data sets.

The second, Data Confidentiality, will build upon current NISS research (funded by NSF's Digital Government program) to develop systems for dissemination of disclosure-limited statistical analyses of confidential Federal data, by focusing on emerging issues and opportunities created by the Internet and other new technologies.

Discussions of a number of problems of a biostatistical and bio-pharmaceutical nature are now taking place with several affiliates, together with a workshop on Analysis of Gene Expression Data, to be held at NISS on July 13, 2000.



Affiliates Kickoff Meeting at NISS:
March 3, 2000

NISS at Interface 2000

NISS had a strong presence at the Interface 2000 meeting in New Orleans in April. A session on *Statistics and Information Technology* featured presentations on three NISS projects in IT.

Todd Graves, ex-NISS postdoc, spoke on "How Should We Publish Journals in the Web Age?" and presented ideas growing out *Live Documents* component of the NISS code decay project (see story on page 4) .

Ashish Sanil, current NISS postdoc, talked on "Geographic Aggregation as a Means of Disclosure Limitation," presenting material from the ongoing digital government project.

Nandini Raghavan, ex-NISS visitor, discussed "Detecting Defectors: Utilizing Massive Data to Predict ISP Customer Churn." This work, in which ex-NISS postdoc Matthias Schonlau (now at RAND) played a major role, was part of the NISS project *Pilot Projects to Explore Large Data Sets*.



Todd Graves



Ashish Sanil



Nandini Raghavan

Affiliated Organizations

Corporations and Government Agencies

Amgen (Thousand Oaks, CA)
AT&T Shannon Laboratory (Florham Park, NJ)
Bureau of Labor Statistics (Washington, DC)
Bureau of Transportation Statistics (Washington, DC)
Census Bureau (Washington, DC)
Environmental Protection Agency (Research Triangle Park, NC)
General Motors (Detroit, MI)
GlaxoWellcome (Research Triangle Park, NC)
Bell Laboratories, Lucent Technologies (Murray Hill, NJ and Naperville, IL)
Mayo Clinic (Rochester, MI)
Merck & Company (West Point, PA)
MetaMetrics (Durham, NC)
National Security Agency (Ft. Meade, MD)
National Center for Education Statistics (Washington, DC)
National Institute of Standards and Technology (Gaithersburg, MD)
Pacific Northwest National Laboratory (Richland, WA)
Pfizer Inc. (Groton, CT)
Procter & Gamble (Cincinnati, OH)
Quintiles Transnational (Research Triangle Park, NC)
Research Triangle Institute (Research Triangle Park, NC)
SAS Institute (Cary, NC)
SmithKline Beecham (Collegeville, PA)
Telcordia Technologies (Morristown, NJ)

University Departments

Carnegie Mellon University (Statistics)
Duke University (Statistics and Decision Sciences)
University of Iowa (Statistics)
Iowa State University (Statistics)
University of Maryland Baltimore County (Mathematics and Statistics)
University of Minnesota (Biostatistics)
North Carolina State University (Statistics)
University of Pennsylvania (Statistics)
Ohio State University (Statistics)
Purdue University (Statistics)
Stanford University (Statistics)
University of North Carolina at Chapel Hill (Biostatistics)
University of North Carolina at Chapel Hill (Statistics)