



National Institute of Statistical Sciences
PO Box 14006, Research Triangle Park, NC 27709-4006
Tel: 919.685.9300 FAX: 919.685.9310
www.niss.org

Annual Report of the Director

Alan F. Karr
October 22, 2002

Contents

1	Highlights	1
2	NISS Research	1
2.1	SAMSI	1
2.2	New Projects	2
2.3	Ongoing Projects	3
2.4	Pending Proposals	3
2.5	Planned Proposals	4
3	Affiliates Program	4
3.1	Expanded Program Benefits	4
3.2	2002 Events	4
3.3	Retention and Recruitment	4
3.4	Affiliate-Related Research	5
3.5	Affiliates Services	5
4	Additional Activities	5
4.1	Communication and Outreach	6
4.2	Sacks Award	6
4.3	Initiatives with Other Research Organizations	6
4.4	Service to the Statistical Sciences Community	6
5	Personnel	6
5.1	Assistant Directors	6
5.2	Research Staff	7
5.3	Postdoctoral Fellows	7
5.4	Staff	7
6	Facilities	7
7	Finances	8
8	NISS Governance	8

1 Highlights

Overall, the past twelve months have been very good for NISS. Highlights of the year include:

Assistant Directors. NISS now has three talented, energetic Assistant Directors: Thomas Gerig, Gary McDonald and Stanley Young. See §5.1 for details.

SAMSI. The Division of Mathematical Sciences at NSF has funded SAMSI as a consortium activity of NISS, Duke, NCSU and UNC. Total funding from NSF is \$10,000,000 over five years. See §2.1.

New NSF Projects. Three other NSF projects have been funded, including a \$2.56M second Digital Government (DG) project. See §2.2.

Affiliates Program. Despite difficult financial times, the program is continuing to develop. Technology Days have become the primary mode of information dissemination to Affiliates. Program benefits have been expanded significantly, in particular to reimburse participation in NISS or SAMSI events.

Finances. NISS finances are more stable than at any point over the past five years. SAMSI and the affiliates program provide a steady source of income, and the new NSF awards range in duration from 2 to 5 years. Sufficient revenue exists to support the additional personnel, while expenditures remain under tight control.

Revised Governance Structure. The NISS Corporation and Board of Trustees were restructured in ways that support greater engagement in NISS affairs by Members and Trustees.

2 NISS Research

2.1 SAMSI

Following a successful site visit in November, 2001, the Division of Mathematical Sciences (DMS) at National Science Foundation (NSF) elected to fund the Statistical and Applied Mathematical Sciences Institute (SAMSI). A partnership of NSF with the consortium of NISS, Duke University (Duke), North Carolina State University (NCSU) and the University of North Carolina at Chapel Hill (UNC), SAMSI is already at work to achieve its vision of forging “new synthesis of the statistical sciences with the applied mathematical sciences and disciplinary science to confront the very hardest and most important data- and model-driven scientific challenges.” SAMSI is led by a directorate of James Berger (Duke), H. T. Banks (NCSU), J. S. Marron (UNC) and me.

SAMSI’s grand opening ceremony took place on September 3, 2002, with more than 200 in attendance. The first two SAMSI programs—*Inverse Problem Methodology In Complex Stochastic Models* (INV) and *Stochastic Computation* (STOCOM)—are underway. A third program, *Large-Scale Computer Models for Environmental Systems* (ENV), will take place in the Spring of 2003.

Interaction between NISS and SAMSI has three principal forms, which in my dual roles as Director of NISS and Associate Director of SAMSI I pursue aggressively:

Scientific. There is strong congruence between the scientific activities of SAMSI and the NISS research program. The 2002–03 INV and ENV programs relate directly to NISS research on computer model evaluation; a 2003–04 program on *Data Mining and Machine Learning* not only relates directly to NISS research on DG and information technology (IT), but also responds to strong interest among the Affiliates.

Visibility. Through its events, which attract large numbers of attendees (Each of the two kickoff workshops had an attendance of more than 120.), SAMSI generates significant visibility for NISS. Events too large to hold at NISS have at least one activity (such as a poster session) at NISS.

Financial. SAMSI is housed on the second floor of the NISS building, for which it pays market rate rent. SAMSI also purchases services such as telephone and computer access from NISS. The two institutes share a computer system manager and communications manager.

2.2 New Projects

In addition to SAMSI, five projects have been funded since January 1, 2002, which are summarized in Table 1. Total funding for these projects exceeds \$3.5 million.

Digital Government (DGII). This project, *Data Confidentiality, Data Quality and Data Integration for Federal Databases: Foundations to Software Prototypes* is a four-year, \$2,560,000 to the first NISS DG project (§2.3). I am PI; Stephen Fienberg of Carnegie Mellon University (CMU) is co-PI. Funding comes from NSF and five Federal agency partners: the Bureau of Labor Statistics, Bureau of Transportation Statistics, Census Bureau, National Agricultural Statistics Service and National Center for Education Statistics. Other participants, whose fields include computer science, the social sciences and statistics, are drawn from the Los Alamos National Laboratory, the University of Maryland, the University of Michigan, Purdue University and Southern Methodist University, as well as the Federal agency partners.

The research addresses a broad set of statistical problems with special application to unique problems faced by Federal government agencies – especially the necessity, in a electronic world, to balance privacy and confidentiality against user access to high-quality statistical data.

ITR–Light Weight Instrumentation (LWI). The goal of this five-year project, entitled *Collaborative Research: Acquiring Accurate Dynamic Field Data Using Lightweight Instrumentation*, is to design, implement and evaluate light–weight instrumentation for software testing and profiling. The underlying concept is that different instances (users’ copies) of the software are instrumented (to collect information) differently but lightly (so that performance is not impaired).

The project is funded by the CISE Directorate of NSF under the Information Technology Research (ITR) program. Ashish Sanil, a yet-to-be-appointed postdoc and I are the NISS personnel. Collaborators include software engineers from the University of Maryland, the University of Washington, the University of California Irvine and Georgia Tech.

ITR–Web (WEB). The thrust of the project (*Bayesian Models Linking Web Site Structure and Usage*) is to relate visitor behavior (page visits) to Web site structure through a series of increasingly complex Bayesian models. Data will be drawn from a variety of types of sites, potentially including an E-tailer, a financial services site, a content site and an information site. Ashish Sanil (who is co-PI), a to-be-appointed postdoc and I are the NISS personnel. Partners include 2nd Sight, Inc. (a provider of Web site analytics) and Visintuit (Stephen Eick, a key participant in the “code decay” project).

Data Quality (DQII). This follow-on to NISS’ first data quality (DQ) project with Bureau of Transportation Statistics (BTS) addresses three issues: (1) An assessment of the state-of-the-art in DQ; (2) An assessment of BTS’ DQ practices; (3) Recommendations for future efforts by BTS. Ashish Sanil and I are the project personnel.

Global Climate Change (GCC). This project funds an FAA postdoc (§3.5) to work on statistical issues associated with global climate change. The specific foci are: 1) Trend analysis of extremes; (2) Modeling extremes in temperature; and (3) Network design for extremes. Amy Grady is the postdoc; Richard Smith (Statistics, UNC) and Gabriele Hegerl (Earth and Ocean Sciences, Duke) are also involved.

Project	Funding Source	Start Date	Duration (Years)	Funds to NISS	Principal Collaborators
DGII	NSF/DG; BLS, BTS, Census, NASS, NCES	8/1/02	4	\$2,560,000	CMU, LANL, Purdue, UMich, SMU
LWI	NSF/ITR	9/1/02	5	500,000	GA Tech, UC Irvine, UMD, UWA
WEB	NSF/ITR	10/1/02	2	235,000	2nd Sight, Visintuit
DQ II	BTS	10/1/02	1	79,000	
GCC	EPA	9/1/02	2.25	257,000	Duke, UNC

Table 1: Summary of New Projects.

2.3 Ongoing Projects

Over the past year, there have been two main threads in the NISS research program.

DG and Associated Activities. The NSF-funded DG project *A Web-Based Query System for Disclosure-Limited Statistical Analysis of Confidential Data* focused on three problem areas:

Tabular data, with innovations ranging from new theory and methodology to an entirely novel dissemination technology known as optimal tabular releases;

Data swapping, with additional funding from the National Center for Education Statistics (NCES), leading to a software system for data swapping and analysis of swapping that has been implemented as both a standalone system and as a Web service.

Regression servers, which are dissemination systems (dynamic or static and optimal) that respond to queries for regressions involving attributes in a database.

Products of the project were presented to the Affiliates at a “Data Confidentiality Technology Day” (§3.2).

Principal participants during the year were Adrian Dobra (NISS), George Duncan (CMU), Stephen Fienberg (CMU), Shanti Gomatam (NISS), Sallie Keller–McNulty (LANL), Stephen Roehrig (CMU), Ashish Sanil (NISS), Lynne Stokes (Southern Methodist) and I.

Computer Model Evaluation. This effort comprises two linked projects: (1) An NSF-funded Focused Research Group (FRG) project *Statistical Framework for Evaluation of Complex Computer Models*; and (2) a targeted project funded by NISS Affiliate General Motors (GM). It is led by James Berger (Duke) and Jerome Sacks (Duke); NISS postdoc Marc Kennedy and visitor Susie Bayarri are also principal contributors. The research addresses Bayesian methodology for model validation, and focused during the year on three testbeds — a stochastic traffic microsimulation, a spot weld model and a model for automobile crash tests.

2.4 Pending Proposals

Two proposals are pending as of the date of this report.

Computational Algebraic Statistics (CAS). This proposal has been submitted to the current FRG competition at NSF. Its goal is to create a new synthesis of computational algebraic geometry and statistics, which will generate theory and methodology applicable to important statistical problems. The research emphasizes problems in large, multi-dimensional contingency tables. The PIs are Stephen Fienberg (CMU), Bernd Sturmfels (UC Berkeley), Jesus De Loera (UC Davis), Serkan Hosten (San Francisco State) and me.

Computer Model Evaluation (GMII). A proposal for a follow-on to the first GM project on computer model evaluation is under development. James Berger (Duke) and Jerome Sacks (Duke) will continue as scientific leaders.

2.5 Planned Proposals

A proposal to the 2003 NSF ITR competition dealing with visualization servers is in the early stages of development. Stanley Young, as Assistant Director for Bioinformatics, is pursuing a number of initiatives in this area, several of are likely to become proposals.

3 Affiliates Program

The original terms of NISS' participating in SAMSI were and remain that SAMSI will not have an affiliates program of its own. However, NISS affiliates are in effect SAMSI affiliates as well (see§3.1). Reflecting the close relationship between NISS and SAMSI and in order to make it straightforward for mathematics departments to become associated with SAMSI and NISS, the name of the university component of the program has been changed to *NISS/SAMSI University Affiliates*. That of the corporation/Federal agency/national laboratory remains *NISS Affiliates*.

3.1 Expanded Program Benefits

In order to make the value added of the Affiliates Program more concrete, effective January 1, 2002, a portion of each affiliate's annual dues (\$2500 of \$10,000 for corporations, Federal agencies and national laboratories, and \$1500 of \$3000 for universities) is placed in an *Affiliates Reimbursement Account*. These funds may be used to reimburse participation in *NISS or SAMSI* events.

In addition, university affiliates memberships may now be shared by two (or more) departments on the same campus (for example, a statistics department and a mathematics department).

3.2 2002 Events

Affiliates events currently consist of two kinds of activities:

Technical Events: To a significant degree, workshops have been replaced by *Technology Days* that give affiliates early access to NISS research and *Short Courses* on important areas of theory and methodology. Workshops through which affiliates inform the course of NISS will continue to be held.

Planning Meetings: Two are held each year, in March at NISS and in August at the JSM.

Affiliates events for 2003 are summarized in Table 2.

3.3 Retention and Recruitment

Summary. With the exception of one corporation (Boeing, because of financial circumstances) and one university department (where NISS' "champion" departed), retention in 2002 was 100%. One earlier affiliate, the Los Alamos National Laboratory, returned to the program.

Event	Date	Location	Attendance
Data Quality Technology Day	2/28/02	NISS	25
Annual Planning Meeting	3/1/02	NISS	30
Data Confidentiality Technology Day	5/29/02	BLS (DC)	35
JSM Affiliates Meeting	8/11/02	New York	35
High Throughput Screening Technology Day	10/25/02	NISS	20
Short Course on Computer Experiments	11/7–8/02	NISS	25

Table 2: Summary of Principal Affiliates Events during 2002.

University Campaign. With the assistance of the Affiliates Committee of the Board of Trustees, a campaign was undertaken in the Spring to 2002, using SAMSI as leverage, to increase the number of university affiliates. More than forty leading statistical sciences departments were approached. Seven new affiliates (Georgia, Maryland, Michigan, Missouri, Oakland University, Southern Methodist and UCLA) were generated as a result. Continuing discussions are under way with several others. One of Thomas Gerig’s responsibilities is to pursue this campaign.

Mathematics Affiliates. An intensive effort is about to begin to attract university mathematics departments with significant activity in applied mathematics or statistics to become NISS/SAMSI University Affiliates. Thomas Gerig and Thomas Banks (Associate Director of SAMSI) are leading this activity.

3.4 Affiliate–Related Research

During 2002, NISS was engaged in research funded by, or in collaboration with the Bureau of Labor Statistics, Bureau of Transportation Statistics, Census Bureau, Environmental Protection Agency, General Motors, GlaxoSmithKline, Los Alamos National Laboratory and National Center for Education Statistics.

3.5 Affiliates Services

Job Listings. A Web-based job listing service for all affiliates was initiated in October, 2001. It continues to operate and is available at www.niss.org/affiliates/ajls.html.

Publicity. The most recent announcement about the program, listing current affiliates and inviting others to join, appeared in the December, 2001 issue of *AMSTAT News*. Another announcement is planned for 2002.

Postdoctoral Program for Federal Agency Affiliates. A joint postdoctoral program of NISS and its Federal Agency Affiliates (FAAs) is in operation. Two appointments have been made during the first year, in conjunction with the Census Bureau and EPA.

Summer Internship Program. The NISS Affiliates Internship Program (NAIP) matches graduate students from NISS university affiliates to internships at corporate, government and university affiliates. NISS’ role is to disseminate job descriptions and then assemble and forward to employers resumes from student applicants. The program functioned modestly but effectively in 2002 and will continue in 2003.

4 Additional Activities

NISS is seeking and undertaking activities other than research that serve the statistical sciences community.

4.1 Communication and Outreach

Largely, these are expansions and refinements of ongoing activities.

NISS Newsletter. The quarterly newsletter is distributed by both mail and the NISS Web site. Katherine Kantner continues as the editor.

Monthly updates from me are sent to the Board of Trustees and Members of the Corporation, with separate updates to the Affiliates.

JSM 2002 Events. Jon Kettenring and I reported on NISS to the Board of Directors of the ASA. (I also reported to the IMS Council, of which I am now a member, at the IMS Annual Meeting in Banff.) More than 35 NISS affiliates attended the August 11 JSM Affiliates Meeting, and more than 150 people came to the NISS Reception on Monday evening, August 12. A NISS-sponsored session “Data Quality: A Statistical Perspective” attracted an audience of approximately 80.

4.2 Sacks Award

The 2002 Jerome Sacks Award for Cross-Disciplinary Research was presented to Max Morris of Iowa State University at the NISS JSM Reception.

4.3 Initiatives with Other Research Organizations

EURANDOM. Plans are progressing to implement a Memorandum of Agreement with EURANDOM to establish and implement sensible forms of interaction. Those under consideration include joint workshops and a postdoctoral exchange program.

4.4 Service to the Statistical Sciences Community

NISS hosts the Web site of the Committee of Presidents of Statistical Societies (COPSS), one of our parent organizations.

A number of Triangle statistical organizations, including the North Carolina chapter of the ASA and the Triangle Econometrics Workshop, now meet regularly at NISS. With the advent of SAMSI, this service to the local community will increase: for example, one SAMSI-related course meets at NISS this Fall, and two will in the Spring of 2003.

At the request of the NCES, in August 2002, NISS conducted a panel to review NCES’ proposed new statistical standards. This is the latest in a series of panel meetings run by NISS for NCES.

5 Personnel

The most important change at NISS over the past year is the appointment of three Assistant Directors.

5.1 Assistant Directors

NISS currently has three part-time Assistant Directors. All have significant responsibilities relative to the Affiliates Program, and two are playing key roles in expanding and diversifying the research program.

Thomas Gerig, Assistant Director for NISS/SAMSI University Affiliates, has as primary responsibility the “care, feeding” and recruitment of NISS/SAMSI University Affiliates. He is continuing and expanding efforts to recruit statistics departments as well as the new campaign (see §3.3) to attract mathematics departments.

Gary McDonald, Assistant Director for Program Development, is playing a major role in expanding NISS research on computer model evaluation (including possible creation of a long-term capability), especially in connection with affiliates. He also plans a significant role in the SAMSI 2003–04 program on Data Mining and Machine Learning.

Stanley Young, as Assistant Director for Bioinformatics, is developing research programs in such areas as drug discovery, genomics and proteomics, as well as interaction with affiliates in the pharmaceutical industry.

5.2 Research Staff

Ashish Sanil is now Research Statistician at NISS. He has a major role in research on DG, DQ and software engineering, and is co-PI on the WEB project (§2.2).

5.3 Postdoctoral Fellows

As of October, 2002, there are four NISS postdoctoral fellows. Three (Shanti Gomatam, Amy Grady and Rui Paulo, who is shared with SAMSI) are resident at NISS; one (John Aston) is located at the Census Bureau under the auspices of the FAA postdoctoral program (§3.5). Two postdoctoral fellows left during the year, one (Adrian Dobra) to join to Duke and the other (Li Liu) to join Pfizer.

Postdoctoral recruiting is of the utmost priority. Based solely on funds in hand, there are work and support for *at least three* additional postdoctorals.

5.4 Staff

The NISS staff consists of three uncommonly dedicated and efficient individuals. **Martha Williamson**, Administrative Assistant, strong support for me and functions as business manager of NISS. **Katherine Kantner** has major responsibilities in connection with the Affiliates Program, the Board of Trustees and communication, including Web sites, for both NISS and SAMSI. **James Thomas**, now full-time computer system manager, provides system, software and network support to NISS and SAMSI.

6 Facilities

With SAMSI’s occupying the second floor of the NISS building, a number of changes have been made to enable more efficient use of space in the building. These include moving the first floor conference room and reconfiguration of the lobby. In addition, the entire interior was painted during the summer of 2002.

Major improvements were made to the first floor classroom, which is shared by NISS and SAMSI. New video origination facilities have been installed that will both benefit NISS and help the community explore the potential of new collaborative technologies.

7 Finances

NISS was financially stable in FY 2002 (ended June 30, 2002) and will be so again in FY 2003.

8 NISS Governance

The *ad hoc* Governance Committee (Stephen Fienberg, John Geweke, Sally Morton and Daniel Solomon, Chair) proposed significant changes to the By-Laws of NISS (Corporation and Board of Trustees), with the goal of maximizing Member and Trustee participation in NISS. The changes were approved by the Board of Trustees in September.

The major changes us reduction of the size of the Board of Trustees from 41 to 21. Other changes: alignment of terms of appointed Members, Trustees and officers with the NISS fiscal year, creation of Evaluation and Compensation and Finance Committees, and abolition of the Scientific Advisory Committee. There were also a number of “housekeeping changes” that recognize current practice.