NISS

Data Confidentiality Technology Day

November 20, 2003
Washington, DC

The Program

9:30 AM	Welcome and Introductions
9:45	Introduction and Overview: Alan Karr
10:00	The NISS Data Swapping Toolkit: Ashish Sanil
10:45	Break
11:00	Releases of Conditionals and Marginals: Aleksandra Slavkovic
11:45	Perturbation Methods: Stephen Roehrig
12:30 P	Lunch
1:30	Regression Servers: Jerome Reiter
2:15	Secure Integrated Analyses: Alan Karr
2:45	Future Plans: Alan Karr
3:00	Adjournment

Introduction and Overview

Thrusts of the DG II Proposal

- DC = Data confidentiality
  - Problem formulations and scalable tools that accommodate both *disclosure risk* and *data/information utility*
- DI = Data integration
  - Understanding *consequences of DI* for DC, DQ and statistical inference
- DQ = Data quality
  - Fundamental quantifications, usable models, scalable methods

What Won't Be Covered

- Record linkage and matching
  - W. Cohen, P. Ravikumar, and S. E. Fienberg (2003). "A comparison of string distance metrics for name-matching tasks." IJCAI 2003 Workshop on Information Integration on the Web.
  - Mikael Bilenko, Ray Mooney, William Cohen, Pradeep Ravikumar & Stephen Fienberg, "Adaptive name-matching in information integration." IEEE Intelligent Systems (2003).
  - W. Cohen, P. Ravikumar and S. E. Fienberg "A comparison of string metrics for matching names and records tasks." To appear in KDD 2003 Workshop on Data Cleaning, Record Linkage, and Object Consolidation.

What Won't Be Covered—2

- Complete risk-utility formulation for data swapping
  - Shanti Gomatam, Alan F. Karr and Ashish P. Sanil, "A risk-utility framework for categorical data swapping." NISS technical report.
  - Shanti Gomatam, Alan F. Karr and Ashish P. Sanil, "Data swapping as a decision problem." Manuscript nearing completion.
- Other work on data swapping
  - Bayesian (prior on swap rate) reconstruction of pre-swap data

What Won’t Be Covered—3

- Continuing work on computational issues for tables
  - Adrian Dobra, Stephen E. Fienberg, Alan F. Karr and Ashish P. Sanil. “Software systems for tabular data releases.” *Int. J. Uncertainty, Fuzziness and Knowledge Based Systems* **10(5)** (2002) 529-544.
  - Adrian Dobra, Alan F. Karr and Ashish P. Sanil, **“**Preserving confidentiality of high-dimensional tabulated data: Statistical and computational issues.” *Statistics and Computing* **13** (2003) 363-370.