Census Data for Transportation Planning—Some Thoughts

George Duncan

Carnegie Mellon
Policy • Management • Information Technology | **Heinz School**

Three Themes

- Confidentiality
- Disclosure limitation methods
- Multiple sources of Information

Confidentiality

- Identified by Jay Waite of Census
- More restrictions? Concern of George Schoener of DOT
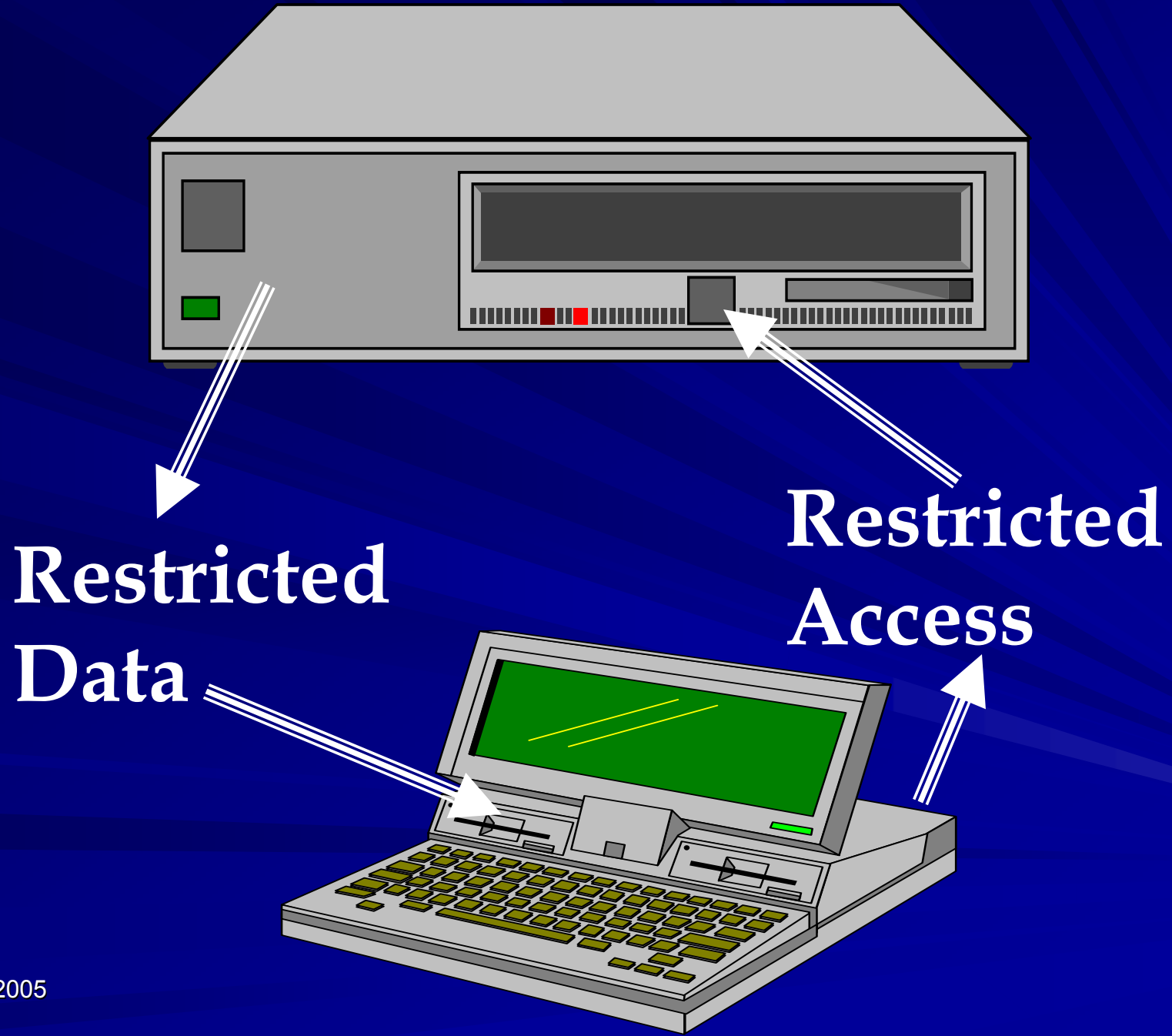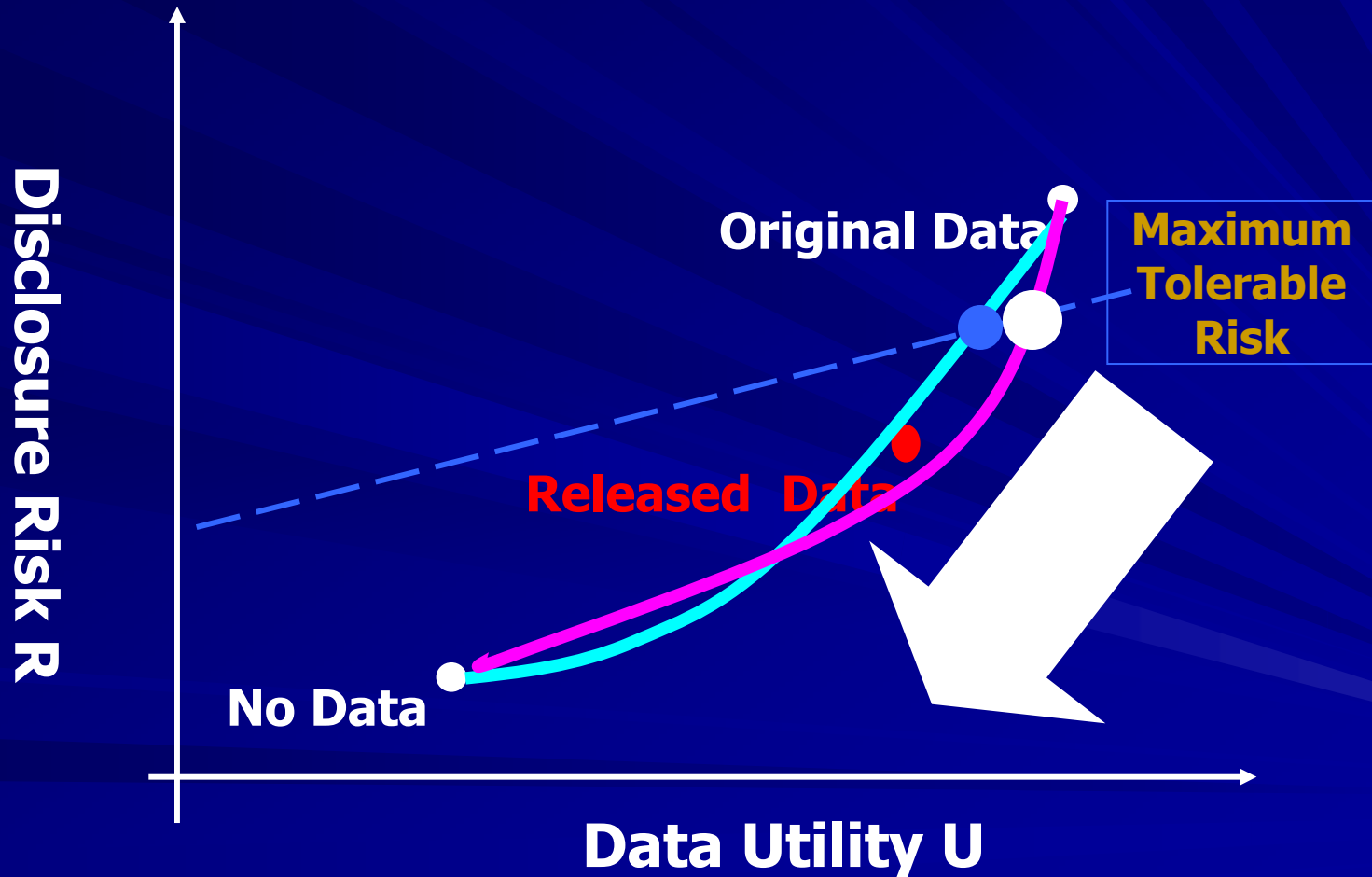- Repeated theme throughout conference

RESPECT THE RATTLESNAKE'S RIGHT TO PRIVACY - PLEASE STAY ON THE TRAIL

PLEASE DO NOT CLIMB WALL

Privacy

- "Privacy is dead, deal with it," Sun MicroSystems CEO Scott McNealy
- But like the rattlesnake, it can bite you …

Why Confidentiality Matters

- Ethical: Keeping promises; basic value tied to privacy concerns of solitude, autonomy and individuality

- Pragmatic: Without confidentiality, respondent may not provide data; worse, may provide inaccurate data
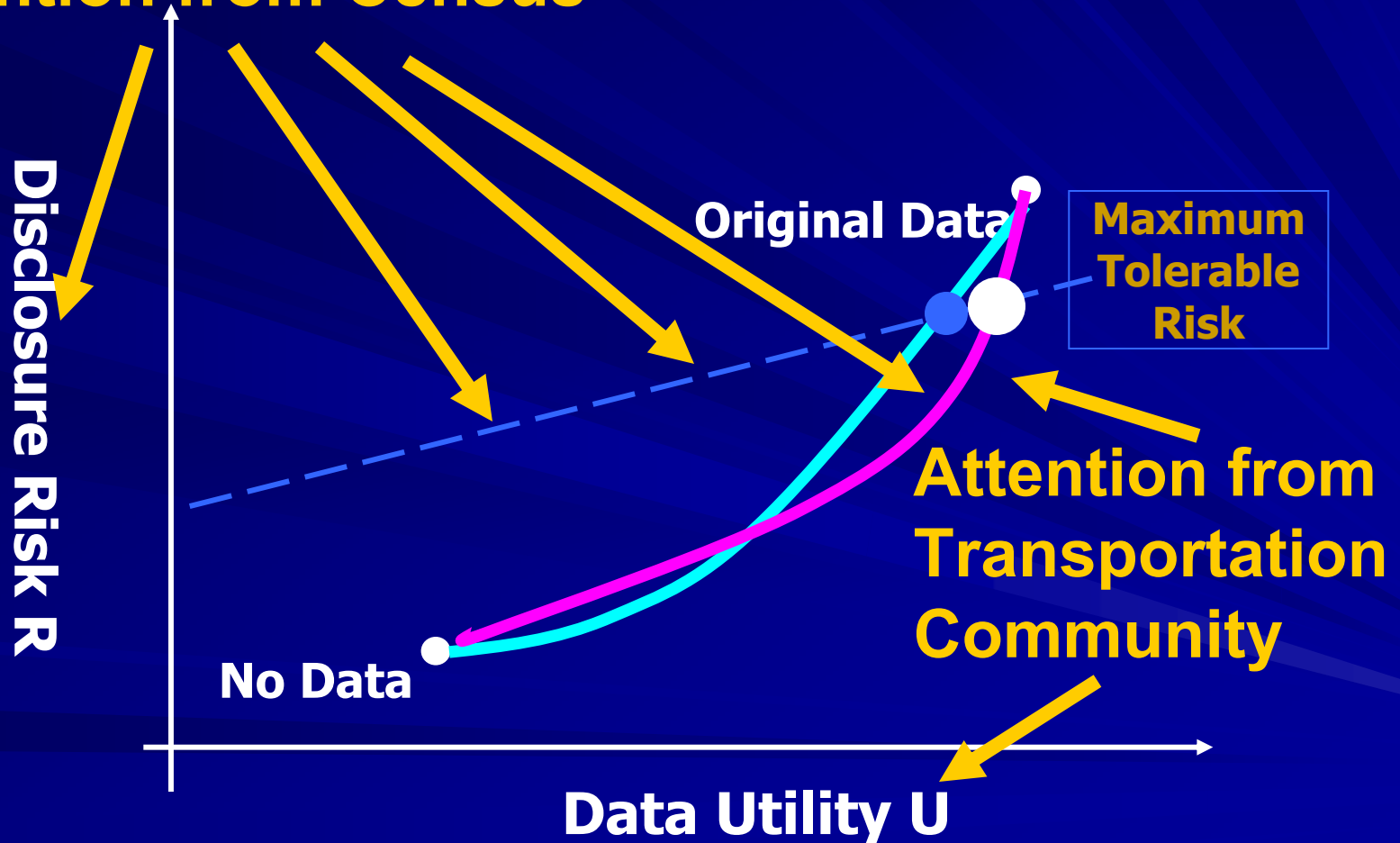
- Legal: Required under law

Restricted Data

Restricted Access

R-U Confidentiality Map

R-U Confidentiality Map

Attention from Census

Disclosure Risk R

Original Data

Maximum Tolerable Risk

Attention from Transportation Community

No Data

Data Utility U

Questions for Joint Consideration

- Does old data decay and so pose less disclosure risk?
- Can providing seasonal data increase data utility without increasing disclosure risk?
- How can the transportation community be educated about the impact of disclosure limitation?
- What specifically causes increased disclosure risk?
- How to improve access to Research Data Centers?

Disclosure Limitation for Tables

- **Coarsening**
  - Aggregate attributes
- **Suppress some cells**
  - Publish only the marginal totals
  - Suppress the sensitive cells, plus others as necessary
- **Perturb some cells**
  - Round
  - Fuzz

Perturbation Methods

- Controlled rounding (Cox)
- Cyclic perturbation (Duncan & Roehrig)
  - Stochastically modify cell values in a known way, allowing a Bayesian analysis of cell value distributions

Multiple Sources of Information

- Alan Pisarski, Thomas Palumbo
- Research and Academic Breakout Group
- Session 8
- "Mixing apples and oranges produces fruit salad"
- "Institutionalize creative quilting" (Nancy McGuckin)

"Institutionalize Creative Quilting"

- **Quilting** is the systematic use of multiple data sources

- **Creative** means drawing on the available resources to address a particular problem

- **Institutionalize** means establishing an apparatus whereby creative quilting can consistently happen

Quilting

- ACS can't provide all transportation planning data needs—national data not be all and end all
- Integrate with other data sets to maximize combined utility to support decision making
- Complement with NHTS, private sector databases, public domain property data, satellite imagery, etc
- Build in dynamic databases that assess changing world—close to real-time data
- Need new methodologies to produce a high-quality data quilt (data stitching)

Creative

- Constrain focus to particular decision making and policy needs
- Search broadly and be imaginative in use of data products

Institutionalize

- Not left to individual planner/researcher
- Identify commonalities and differences among data users, say MPOs
- Find representation for users with common needs, say AASHTO
- Establish ongoing links with information organizations, say Census

Scenario of Potential

- Several MPOs have similar data needs
- AASHTO solicits data needs and formulates a "quilt" to cover them
- AASHTO negotiates with Census for
  - way to develop disclosure limited data product
  - Access to Census Research Data Center to validate statistical inferences from original data ("red light"/"green light")

Opportunities and Threats

- **Opportunity:** Together, Census and Transportation support respective institutional needs for confidentiality and access to quality data

- **Threat:** Divided, Census and Transportation fall to privacy appeals or inability to cope with methodological challenges of disclosure limitation and complex, new data sources