

# Databases must balance privacy, utility, says statistics professor

August 30 2007

---

Agencies like the U.S. Census Bureau produce a voluminous amount of data, much of which is of tremendous value to social scientists and other researchers. But the data also includes personal information that, under the law, must be protected and could be harmful were it to fall into the wrong hands.

Thus, organizations that maintain such databases need to devise ways to protect individuals' privacy while preserving the value of the information to researchers, writes Carnegie Mellon University Statistics Professor George Duncan in a commentary in the Aug. 31 edition of the journal *Science*.

Duncan said traditional methods of "de-identifying" records, such as stripping away Social Security numbers or birthdates, are inadequate to safeguard privacy because a person who knows enough about the data pool could use other characteristics to identify individuals. Duncan, for example, is the only person who holds a Ph.D. in statistics and teaches in Carnegie Mellon's H. John Heinz III School of Public Policy and Management, so any data set that included that information, even with Duncan's name removed, could be used to determine his identity.

This could have serious consequences when it comes to data that includes information about a person's medical history or sexual behavior, like that collected by the National Center for Health Statistics. Unfortunately, the characteristics that can be used to re-identify records are often the very information that makes the data useful to legitimate

researchers.

“The question is, ‘How can data be made useful for research purposes without compromising the confidentiality of those who provided the data?’” Duncan said.

Possible solutions to this dilemma include administrative procedures that limit data access to approved users who must abide by restrictions on the use of information, and statistical methods that de-identify records in such a way that the user cannot readily reconstruct personal identities. In order to be effective, these statistical transformations must be tailored to how the data will be used, so that researchers can see the information that interests them while other characteristics remain veiled.

Duncan’s commentary in *Science* was prompted by recent reports on data privacy, one by the U.S. National Research Council and the other by the U.K. Royal Academy of Engineering. In the article, Duncan discusses efforts to safeguard information gathered by video surveillance cameras, wireless networks and radio-frequency identification tags, which are used by hospitals to ensure that patients receive the correct treatment.

“Achieving ‘adequate’ privacy will require engineering innovation, managerial commitment, information cooperation of data subjects and social controls (legislation, regulation, codes of conduct by professional associations and response to reactions of the public),” Duncan wrote.

Source: Carnegie Mellon University

Citation: Databases must balance privacy, utility, says statistics professor (2007, August 30)  
retrieved 23 April 2024 from

<https://phys.org/news/2007-08-databases-privacy-statistics-professor.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.