

Endorsement by the American Statistical Association of the Statement on Intellectual Property Protection for Databases by the American Association for the Advancement of Science

At its December 6, 1997 meeting, the American Statistical Association (ASA) Board of Directors voted unanimously to endorse the enclosed Statement on Intellectual Property Protection for Databases from the American Association for the Advancement of Science (AAAS). We appreciate the leadership shown by AAAS and by the National Research Council in efforts to protect the Constitutional balance of respect both for the public's right to knowledge and for the rights of intellectual property owners in the face of advancing information technologies.

In addition to the principles laid out in the AAAS statement, ASA feels compelled to emphasize a potential unintended danger to the right of independent statistical analysis. Under both "investment protection" and "sui generis" principles of database protection, proposed treaties and laws have sought to safeguard the economic interests of makers or owners of databases beyond the mere publication of single database products. They would protect the owners' database rights in both actual and potential markets perpetually. There is no evidence of any intent to preclude or restrict the rights of statisticians to analyze databases, which in their raw form might be quite daunting to review and analyze.

Still, it is clear that the publication or other dissemination of a competent statistical analysis or summary of the major information content of large databases could, in some cases, reduce the market scope for the raw data itself. This would not be true of databases where the major value comes from the accessibility of individual items, such as physicians' specialties and addresses in a physician directory.

An example of an analysis potentially harming a database owner's economic interest is as follows. Suppose that a database publisher has compiled nonproprietary geographic data on the distributions of family size and income relative to distance from various commercial locations for a metropolitan area with the intent of supporting marketing analysis by various commercial businesses. Assume that the database consists of long tables of accurate records showing the locations and incomes of families and the locations of commercial centers. There is a query system with it, but no summary treatment of the data. Suppose further that a statistician with legitimate access to the database turns this into one or more map-based spatial diagrams of densities of commercial activity and densities of families grouped by income. The motivation to do so might be either scholarly or commercial.

Marketing researchers for businesses in any of the commercial centers could tell instantly by looking at these diagrams the relative distances from their locations of various income clusters of families. Factoring in local knowledge of the transportation corridors, they could quickly assess the effort it would take to appeal to potential customers of different income levels. If they do not already own the original database, are they likely to buy it? If they bought it, they would have to figure out how to analyze it. The statistician's product is likely both to be much cheaper than the database and to be directly interpretable without further number-crunching. Hence, the market for the database would decrease. (It would still be bought by those who are unaware of the

statistical product or who feel they need the detailed data or who are not convinced that the summary is accurate.)

The statistical product is not strictly a value-added product because it does not contain the original data. (It would be piracy if it did contain it without permission.) It is not a database product at all. It may not even be directed at the market for the original data. In spite of all of these factors which would appear to support its innocence, it does threaten the economic value of the original database.

Under proposed regimes, the database maker might be able to sue and win. In a court challenge, a judge could rule that the statistician was in violation of law for damaging the economic rights of the database owner. Should this occur, the common, valuable, and well-respected practice of statistical analysis could be "chilled" by the threat of lawsuits.

This example was constructed for general public understanding and for its statistical character. The danger it represents, however, potential restriction or chilling of scientifically based analysis and publication, threatens the work of scientists and engineers more generally. Science builds on the seminal work of earlier authors. If owners of scientific databases were to have exclusive rights to analyze and publish information derivable from their databases, scientific progress would be greatly restricted. If the owners of databases of structural materials properties were to have the exclusive rights to apply those properties to the analysis and design of structures, engineering innovation would be seriously hindered.

In science and engineering, when database owners wish to protect their analytic rights, they delay publication of the raw data until they can publish their own analyses. Once the data have been released, anyone is free to perform competing or alternative analyses of the same data. This practice is essential to science and to public input into policy analysis and decision making. The right to perform statistical (or other scientific and engineering) analyses must not be abridged by database protection regimes.

The ASA has another concern, as well. Proposed new database protection regimes properly tend to exempt government data from access restrictions. It should also be recognized that the private sector in some cases contributes databases to the public domain under principles of full and open access to factual data - either unilaterally or during the course of work in government-industry partnerships. The latter might be termed "government-equivalent" data because of its contribution to the freely accessible public store of knowledge. Neither government data nor privately donated government-equivalent data should be subject to restrictions on reproduction or usage - either in original dissemination or through incorporation into subsequent value-added products of database entrepreneurs.