CGTM 74
Lance J. Hoffman
August 1969

Social Data vs. Individual Privacy: Computers, Privacy, and the National Data Center

A panel discussion at the 1969 Meetings of the American Psychological Association in Washington D.C.

Introductory Remarks by

Lance J. Hoffman
Stanford Linear Accelerator Center
Computation Group
Stanford University
Stanford, California 94305

One of the most notable disadvantages of a central facility for data storage and retrieval is the threat to individual privacy inherent in a centralized data bank. By centralizing information all in one place, we achieve economies of scale for both the legitimate and the illegitimate users of the facility. An unauthorized disclosure of information from a large, central data bank can potentially be much more damaging than one from a smaller data bank. It has been argued that the National Data Center would be a statistical data bank, not an intelligence, or dossier, data bank. From a technological point of view, this distinction is baseless. Any statistical data bank can be used as an intelligence data bank and vice versa.

Social scientists can incorporate certain safeguards into computer systems they use which will tend to lessen the threats to individual privacy.

Let me mention some:

1. On request, any person should be able to see a printout of all factual data relating to him which the data bank contains. He should be notified periodically of the most recent uses of data relating to him, including who received or put in the data.

2. Minimal encoding of the data should be performed. No so-called "clear" information should be stored in the data bank. The technological cost of doing this is very small and no expertise is required - one method has already been published in an "underground" newspaper. Casual snooping can be easily prevented in this way.
3. Obvious identifying information such as names and social security numbers should not be kept with the data itself.
4. When possible, samples such as a 1 in 1000 sample should be taken of populations (rather than polling the entire population). This would minimize the chance of personal data about any one particular person being in the system.
5. Care should be taken to see that statistical output which is released describes samples which are sufficiently large to block possibilities of identifying specific data by deduction.
6. Data inoculation techniques should be developed to insure that even if a record is associated with a particular individual, there is no guarantee that the record is entirely accurate. These techniques can induce errors which are sufficiently small so that the statistical findings will still be valid, yet the data on any given individual may have been altered to be partially or entirely wrong.

Unless both the data processing community and the social sciences begin to devise protective measures for individual data in data banks, they may expect regulations to be imposed on them by federal and local authorities to control this use. These regulations may unnecessarily inhibit computer-aided research and decision-making. Social scientists owe it to themselves to take steps now to reduce the fears of invasion of personal privacy by computer data banks. They can do this by using minimal privacy safeguards (such as those I've described) whenever they use the computer in their work.