

PRIVACY AND EFFICIENT GOVERNMENT: PROPOSALS FOR A NATIONAL DATA CENTER

To an increasing extent, administration of modern government and effective program planning have required the collection and processing of a vast amount of information. Fortunately, the expansion in data handled by governmental agencies has been accompanied by a technological revolution in data processing¹ which promises to make manageable the task of dealing with this avalanche of information. The core of the new technology is the electronic computer and an assortment of ancillary data processing equipment and techniques. The familiar IBM punch card and magnetic data tapes used for storage are already in wide use both in government and in private industry.² Until now the use of this new technology within the federal government has been mostly by individual agencies seeking to meet their individual needs. However, proposals have been made for the creation of a national data center which could take fuller advantage of modern technology by centralizing the federal statistical network. These proposals have been considered by several congressional committees³ concerned about the implications of such a computerized centralization of data for the continued enjoyment of personal privacy. General public reaction to the idea of a national data center has shown a similar fear that personal privacy might be diminished.⁴ Most commentators agree in principle that the efficiency of governmental operations will be increased by the creation of a data center,⁵ but even data center proponents

¹ See A. WESTIN, *PRIVACY AND FREEDOM* 158-68 (1967) [hereinafter cited as WESTIN].

² In 1966 the federal government was using some 2600 separate computers representing an investment of about six billion dollars. *Id.* at 160-61.

³ See *Hearings on Computer Privacy Before the Subcomm. on Admin. Practice and Procedure of the Senate Comm. on the Judiciary*, 90th Cong., 1st Sess. (1967) [hereinafter cited as *Senate Hearings*]; *Hearings on the Computer and Invasion of Privacy Before a Subcomm. of the House Comm. on Gov't Operations*, 89th Cong., 2d Sess. (1966) [hereinafter cited as *House Hearings*]. For a more receptive committee's views see SUBCOMM. ON ECONOMIC STATISTICS OF THE JOINT ECONOMIC COMM., 90th CONG., 1ST SESS., *REPORT ON THE COORDINATION AND INTEGRATION OF GOV'T STATISTICAL PROGRAMS* (Joint Comm. Print 1967) (endorsing the idea of a national data center); *Hearings on the Coordination and Integration of Gov't Statistical Programs Before the Subcomm. on Economic Statistics of the Joint Economic Comm.*, 90th Cong., 1st Sess. (1967) [hereinafter cited as *Joint Hearings*].

⁴ See, e.g., *A Government Watch on 200 Million Americans?*, U.S. NEWS & WORLD REPORT, May 16, 1966, at 56; Hirsch, *Data Banks: The Punchcard Snoopers*, 205 NATION 369 (1967); Miller, *The National Data Center and Personal Privacy*, ATLANTIC, Nov. 1967, at 53.

⁵ See, e.g., *House Hearings* 2 (remarks of Rep. Gallagher), 121 (statement by

recognize the substantial danger that this new efficiency would seriously diminish privacy.⁶

The most recent and comprehensive model for a federal data center is that developed for the Bureau of the Budget by a special task force headed by Professor Carl Kaysen. This Note will inquire, in the light of the Kaysen model and other conceivable models, whether American society can benefit fully from the increased efficiency of a computerized data center without a significant sacrifice in personal privacy and individual liberty.⁷

I. THE KAYSEN TASK FORCE REPORT

The motivating force behind current proposals for a national data center seems to be the concern of social scientists over the inaccessibility for scholarly research of the wealth of data held by the federal government. A three-year study completed by the Social Science Research Council (SSRC) of the American Economic Association in 1965⁸ found that neither outside scholars nor federal agencies were able to utilize public data efficiently because of excessive decentralization in the maintenance of the data files. Moreover, the pressures of ordinary business within individual agencies were found to render the agencies generally unresponsive to outside inquiries. Accordingly, the SSRC urged that a federal data center be established "to preserve and make available to both Federal agencies and non-Government users basic statistical data originating in all Federal agencies."⁹

This recommendation prompted an evaluation by the Bureau of the Budget of the utility of a national data center. A consultant's report called for immediate creation of a data service center.¹⁰ The report estimated that the initial input for the data

Paul Baran, RAND computer expert); Miller, *supra* note 4, at 53-54; *Senate Hearings* 44 (statement of Charles Zwick, Ass't Director, Bureau of the Budget).

⁶ See, e.g., REPORT OF THE TASK FORCE ON THE STORAGE OF AND ACCESS TO GOV'T STATISTICS, ANNEX (1966) (available from the Bureau of the Budget), reprinted in *Senate Hearings* 25 [hereinafter cited as KAYSEN REPORT].

⁷ Similar problems arise in connection with the data collection processes of private organizations. The analysis of this Note is limited, however, to the questions raised by data processing and collection by the federal government as they relate to the proposal for a national data bank. For a discussion of the impact of private data collection upon privacy, see V. PACKARD, *THE NAKED SOCIETY* (1964); M. BRENTON, *THE PRIVACY INVADERS* (1964).

⁸ COMM. ON THE PRESERVATION AND USE OF ECONOMIC DATA, REPORT TO THE SOCIAL SCIENCE RESEARCH COUNCIL OF THE AMERICAN ECONOMIC ASSOCIATION (1965), reprinted in *House Hearings* 195.

⁹ *Id.* at 202.

¹⁰ E. DUNN, STATISTICAL EVALUATION REP. NO. 6—REVIEW OF PROPOSAL FOR A NATIONAL DATA CENTER, reprinted in *House Hearings* 254. Mr. Dunn added the term "service" to his recommendation for a national data center in order to emphasize the need for a central facility that would do more than merely store data

center could consist of about 9000 reels containing representative data principally from the following agencies: the Bureau of the Census (current population and housing data), the Bureau of Labor Statistics, the Internal Revenue Service, and the Social Security Administration.¹¹

The Kaysen Task Force¹² was asked to consider appropriate "measures which should be taken to improve the storage of and the access to"¹³ federal statistics. In its 1966 report, however, the Task Force determined that questions of storage and access could properly be answered only in the context of a broad discussion of how a federal statistical system should operate to satisfy basic statistical needs. These needs include capacity to meet the increasing demand for statistical data, development of safeguards for preserving the privacy of personal disclosures to the Government, maximum utilization of existing data, and minimization of the burden upon citizens and institutions called on to furnish information.¹⁴

The Task Force found the present statistical system inadequate and inefficient in meeting these basic needs.¹⁵ One dimension of the inadequacy is the extended time lag between the receipt of data by an agency and its availability to others in usable form.¹⁶ Individual agencies cannot be relied on to take special measures to make their data more accessible since their primary concern properly is with the application of their limited resources to their specific statutory responsibilities.¹⁷ Another dimension of the inadequacy of currently available information is its publication in merely summary form, which results in the suppression of micro-information and thus makes new analysis difficult and very costly.¹⁸

— it must have a servicing capability, power to establish standards and to monitor compliance. *Id.* at 267.

¹¹ *Id.*, App. B at 276-77. The cost of this initial capability was estimated at \$260,000. This estimate includes only those costs necessary to make data accessible within the responsible agency. Costs of tape copying and blank reels would add about \$500,000 to the total cost. A complete selective data file of about 20,000 reels would take 3 to 5 years to create at a cost of 3 to 3.5 million dollars. *Id.*

¹² The Task Force was composed of Carl Kaysen, chairman (Institute for Advanced Study); Charles C. Holt (Univ. of Wisconsin); Richard Holton (Univ. of California, Berkeley); George Kozmetsky (Univ. of Texas); H. Russell Morrison (Standard Statistics Co.); and Richard Ruggles (Yale University).

¹³ KAYSEN REPORT 1.

¹⁴ *Id.* at 1-2.

¹⁵ *Id.* at 5.

¹⁶ For example, the IRS "Statistics of Income for Corporation Income Tax Returns" is not available in detailed form until 2½ years after the filing of the returns. *Id.*

¹⁷ *Id.* at 13.

¹⁸ *Id.* at 6, 9. For an example of the kind of summary publication now used

The Task Force found present statistical operations inefficient in failing both to keep costs to a minimum and to utilize fully the potential statistical resources.¹⁹ The lack of economy arises from the lack of effective coordination in collecting data and from the operation of many agencies on a scale too small for efficient use of modern cost-saving techniques.²⁰ Low utilization of present potential is the result of practical restrictions on access to the summary publications of data and of incompatibility of the basic unit definitions, classification systems, and techniques of analysis used by different agencies.²¹ The Task Force also found a further restriction on full use of data due to uneven and sometimes excessive application of confidentiality restrictions.²²

To correct these shortcomings the Task Force would create a "model" statistical system, completely centralized in all its essential functions of collection, storage, and analysis of "general purpose" data "not produced as a by-product of the administrative operations of the Government."²³ However, given the present decentralized structure, the Task Force found such a radical change unrealistic and recommended instead that only the storage function be centralized at this time.²⁴ This function was chosen because it was thought to be the most inadequately performed and the most easily separated from the present framework.²⁵ Initially the center would assemble "in a single facility all large-scale systematic bodies of demographic, economic, and social data generated by the present [federal] data-collection or administrative processes."²⁶ In so doing the goal would be to achieve maximum integration of each data record and to provide ready access "within the laws governing disclosure" to all government users. Where appropriate, qualified outside users could benefit from the center on a compensatory basis. The center would cooperate with state and local agencies to achieve even fuller integration of data and increased access to available in-

see, e.g., BUREAU OF THE CENSUS, DEPT OF COMMERCE, CONGRESSIONAL DISTRICT DATA BOOK (DISTRICTS OF THE 89TH CONGRESS) (1966).

¹⁹ KAYSER REPORT 5.

²⁰ *Id.* at 10.

²¹ Another important weakness in the present statistical operation is the absence of adequate file documentation—that is, information about the contents, whereabouts, and meaning of the data. Interview with Paul Kruegar, Ass't Director for Statistical Standards, Bureau of the Budget, in Washington, D.C., June 20, 1968 [hereinafter cited as Kruegar Interview].

²² KAYSER REPORT 9-10.

²³ *Id.* at 13. Individual agencies could make occasional special-purpose studies as well as collect and process data needed in program administration.

²⁴ *Id.* at 16.

²⁵ *Id.* at 17.

²⁶ *Id.*

formation.²⁷ An important function of the center would be establishment and enforcement of uniform standards for closure so that confidentiality could be preserved without present loss of analytically useful information. The Task Force insisted that freer access would not endanger personal privacy but rather that the center's supervision could increase protection of confidentiality.

The proposed data center would be organized within the Executive Office of the President and would be under the control of the "Director of the Federal Statistical System."²⁸ The Director would have two advisory councils: one to represent the interests of government users, the other to speak for the private users and the public-at-large. The councils would advise the Director on such matters as confidentiality, user needs, and the burden on those providing information. The Office of Statistical Standards²⁹ would be transferred from the Bureau of the Budget to become a staff office for the Director. The Bureau of the Budget would also be placed under his control on a coordinate basis with the new data center.

II. THE CONCEPT OF A DATA CENTER

Most criticism of the Kaysen proposal proceeds from the expectation that the center's functions would grow beyond present proposed limitations and that such growth would present intolerable dangers to privacy. An assessment of the Kaysen Task Force proposal depends, therefore, on an analysis of the development which the proposed data center might take in the face of the foreseeable pressures for such development.

The essence of a data center is not the physical concentration of data but rather the capacity to provide particular data or combinations of data upon request. Thus, the simplest data center would be one where a single official had the authority to secure data from participating federal agencies. If agency cooperation

²⁷ *Id.* at 18. Some data exchange programs are already underway. See, e.g., Treas. Reg. § 301.6103(a)-1(d)(1) (1961) (inspection of estate and gift tax returns by states on reciprocal cooperation basis).

²⁸ This would be a new position, filled by presidential appointment, which would assume the coordinating powers over federal statistical programs now assigned to the Office of Statistical Standards of the Bureau of the Budget. See note 29 *infra*.

²⁹ This office exercises certain coordinating functions conferred upon the President by 31 U.S.C. § 18a-b (1964). Although the Bureau of the Budget has assured Congress that it will offer legislation to implement the data center should it finally decide to propose it, see *Senate Hearings* 37 (testimony of Charles Zwick, Assistant Director, Bureau of the Budget), this present statutory power seems sufficiently broad to authorize at least some of the proposed centralization.

were v
and dat
integrat
would
ception
more ef
ected r
collecti
Met
efficien
fied to
mission
unless
nomica
this fu
the ag
sponde
agenc
ing for
in som
warrant
Ce
— a f
centra
vision
ficatio
forme
source
level
Ce
exten
reser
Repo
very
tabul
to qu
A
tical
func
on Ce
for th
conta
of thi

enter would be the standards for dis served without t n. The Task For r personal priva uld increase act

ganized within be under the cont tem." 28 The Dir to represent the eak for the priv s would advise user needs, and Office of Statist Bureau of the Bud Bureau of the a coordinate

ENTER

proceeds from d grow beyond rowth would c ent of the Kay an analysis of ter might take ment.

ysical concentr ticular data or implest data c authority to s agency cooper

dy underway. See state and gift tax rel

tial appointment, whi stical programs now of the Budget. See

onferred upon the F f the Budget has a center should it of Charles Zwick, power seems suffe zation.

were voluntary, the center's access to data might be restricted and data files remain fragmented. Since full access and maximum integration are the goals of an efficient data operation, the center would want unimpeded access to all federal data, with the exception of classified information. The center would be even more effective if it had authority, or perhaps even a legally protected right, to obtain access to state, municipal, and private data collections.³⁰

Mere retrieval capacity, however, cannot provide maximum efficiency. The costs of central storage facilities would seem justified to avoid the higher costs of duplicative requests and transmissions of data from the participating agencies. Moreover, unless the collection of the data could be performed more economically by individual agencies, the data center might assume this function as well. It would receive requests for data from the agencies and prepare questionnaires to send to various respondents or devise and carry out, in consultation with requesting agencies, other plans for obtaining information. Separate reporting for income tax or other purposes might be found unnecessary in some cases. In others, a balancing of the efficiencies might warrant decentralized collection of data.

Central collection would ensure maximum data compatibility—a primary objective of the efficiency interest. Even without central collection the data center should exercise sufficient supervision over data collection to achieve uniform definitions, classifications, and sampling procedures. This task is currently performed to some extent by the Bureau of the Budget, but its resources and effort have been too small to achieve a satisfactory level of coordination.³¹

Centralized collection, storage, and analysis coupled with extensive central supervision over independent data operations resembles the "model" statistical system outlined in the Kaysen Report. As a "statistical" center it would be designed to function very much like the present Bureau of the Census: it would present tabulations, correlation matrices, and similar analyses in response to questions calling for an answer in terms of aggregated data.

Although the Kaysen model was addressed strictly to statistical uses of data, a data center could also serve nonstatistical functions such as tax collection or criminal law enforcement.³²

³⁰ See note 27 *supra*. Constitutional authority for this access could be grounded on Congress' general power of investigation under the necessary and proper clause.

³¹ Kruegar Interview, *supra* note 21.

³² The FBI and the IRS presently use computerized data centers of their own for these purposes. These data centers, used in administering specific programs and containing investigative data, raise special problems not directly within the scope of this Note.

of IRS comment in Levy article

To perform these functions comprehensively and with full efficiency, the data center would have to be capable of nonstatistical or individualized output. Equipped for such output, it could serve as a "dossier center," an intelligence network making on request analyses of individual persons or other units. While able to make statistical analyses, such a center would have the further capability to produce lists of persons associated with specified characteristics. Although a dossier center would be valuable for such tasks as criminal law enforcement, loyalty and security clearances, and personnel screenings, it also would facilitate routine tasks such as addressing envelopes to persons qualified for a particular government program.

The centralized possession of all existing data in a facility capable of individualized as well as statistical output would surely reduce the costs and increase the efficacy of the Government's policing and program administration. A substantial dossier capacity might justify giving the data center independent investigative powers to enable it to make each individual's data record sufficiently complete and accurate to meet all governmental purposes. If, on the other hand, the center is viewed as merely a means of making governmental statistics more accessible and useful for policy planning and scholarly research, a dossier capacity would be unnecessary. Thus, the degree of dossier capacity given to the center and its permitted use of individualized output must turn in part on the services the center is to perform. The Kaysen Task Force seems to have envisioned a relatively narrow range of services.³³ But as the difficulty and complexity of the task of governing the United States increases in future years, and as advances in technology make possible more efficient methods of carrying out this task, it may be expected that pressure would grow for an expansion of the functions of a national data center along the lines suggested above.

III. CONFLICT WITH PRIVACY

A. The Privacy Interest

The phrase "right to privacy" has application in many contexts because it does not refer to a single social interest, but rather subsumes a complex and interrelated set of interests.³⁴

³³ The *Kaysen Report* does not explicitly reject the possibility of individualized output. Later interpretation of the proposal has indicated, however, that the center as proposed would not produce individualized output. See *Senate Hearings*, 14 (testimony of Carl Kaysen), 44 (statement of Charles Zwick, Ass't Director, Bureau of the Budget). Mr. Zwick's statement excepted from the ban on individualized output any disclosures of data already a matter of public record.

³⁴ See generally WESTIN, *supra* note 1, at 8-64.

Privacy is most commonly thought of as the set of interests in secrecy, physical integrity of the person,³⁵ and seclusion. Unrestrained searches and seizures, surreptitious eavesdropping and wiretapping, compelled submissions to lie detector tests and truth serum all violate the interest in privacy so conceived. Since these methods of obtaining information about a man are used without his knowledge or against his will, they violate his privacy even if the information obtained is not such that he would have particularly wanted it kept secret. This range of interests in privacy has in recent years received increasing constitutional protection against the most common invasions.³⁶

But privacy is much more than mere secrecy, physical integrity and seclusion: it is a "very special kind of independence" ³⁷ permitting men to pursue other ends and itself an essential ingredient in many specific values.³⁸ Much of what men are at liberty to do they seek to do in varying degrees of privacy — indeed, without the right to privacy, an activity may lose its significance, its utility, or its joy. For example, in American society a man has the right to think whatever he pleases,³⁹ but if there were public identification of an individual with all his expressions of his thoughts, social pressures might seriously diminish opportunities and incentives for discussion and independent thought.⁴⁰ For this reason a limited constitutional right

tort law an individual's "interest in the integrity of his person includes the things which are in contact or connected with it." W. PROSSER, LAW OF TORTS (3d ed. 1964).

e.g., *Katz v. United States*, 389 U.S. 347 (1967) (nontrespassory eavesdropping is subject to fourth amendment standards); *Camara v. Municipal Court*, 387 U.S. 523 (1967) (administrative searches are subject to fourth amendment standards); *Schmerber v. California*, 384 U.S. 757, 764 (1966) (dictum) (compelled submission to lie detector test would violate privilege against self-incrimination); *Griswold v. Connecticut*, 381 U.S. 479 (1965) (zones of privacy created by first, third, fourth, fifth, and ninth amendments include protection of marital intimacy from invasion by state to enforce ban on use of birth control where state could not show overriding interest in prohibiting use of contraception). *But see* *Schmerber v. California*, *supra* (compelled blood test permis-

W. Prosser, *The Pattern of Liberty*, in ASPECTS OF LIBERTY 15, 17 (M. Konvitz ed., 1958).

generally Fried, *Privacy*, 77 YALE L.J. 475 (1968).
Cantwell v. Connecticut, 310 U.S. 296, 304 (1940) ("freedom to believe is absolute [unlike the freedom to act]"); *West Virginia State Bd. of Educ. v. Barnette*, 319 U.S. 624 (1943) (compulsory flag salute violates first amendment).
 In the *Barnette* case the Court observed that:

... no official, high or petty, can prescribe what shall be orthodox in politics, nationalism, religion or other matters of opinion or force citizens to confess by word or act their faith therein.

at 642.
⁴⁰ The man who is compelled to live every minute of his life among others

UNIVERSITY OF MICHIGAN LIBRARY

to privacy of association⁴¹ and anonymity in expression⁴² has been recognized as incident to freedom of association and of speech.

Moreover, certain aspects of a person's life, such as his deepest fears and hopes, are his exclusive possession. Other aspects he shares with widening circles of companionship — spouse, family, friends, colleagues, strangers — with the intention and expectation that such aspects will not become more widely known. This expectation is sometimes enforceable by a tort action for invasion of privacy,⁴³ but in most cases it is preserved only by tacit understandings that one's confidants will respect social mores governing the propriety of disclosing the content of particular communications. Indeed, the opportunities men have for relative privacy in ordering their affairs and the extent to which their expectations of limited communication are honored may be said to form a measure of their society's regard for individual human dignity.⁴⁴

For most men privacy can be achieved and maintained not by seclusion on a desert island but by existence in a crowded society with the freedom to share with others, or to withhold, their personality — their attitudes, beliefs, opinions, affections, habits, idiosyncracies, and other behavior — in the unfettered exercise of a free will.⁴⁵ Every compelled exposure of this otherwise more or less "private" self represents a loss of privacy. On the other hand, voluntary disclosure to a few intimates, to many friends and associates, or even to the whole society, is not necessarily a loss but may be an expression of the privacy interest.

and whose every need, thought, desire, fancy or gratification is subject to public scrutiny, has been deprived of his individuality and human dignity. Such an individual merges with the mass. His opinions, being public, tend never to be different; his aspirations, being known, tend always to be conventionally accepted ones; his feelings, being openly exhibited, tend to lose their quality of unique personal warmth and to become the feelings of every man. Such a being, although sentient, is fungible; he is not an individual.

Bloustein, *Privacy as an Aspect of Human Dignity: An Answer to Dean Prosser*, 39 N.Y.U.L. REV. 962, 1003 (1964).

⁴¹ See, e.g., *Gibson v. Florida Legislative Investigation Comm.*, 372 U.S. 539 (1963); *Shelton v. Tucker*, 364 U.S. 479 (1960); *NAACP v. Alabama ex rel Patterson*, 357 U.S. 449 (1958). But see *Barenblatt v. United States*, 360 U.S. 109 (1959).

⁴² See *Talley v. California*, 362 U.S. 60 (1960) (invalidating ordinance requiring leaflets to carry name and address of sponsor).

⁴³ At least 30 states and the District of Columbia now recognize a cause of action for invasion of privacy. W. PROSSER, *LAW OF TORTS* 831-32 (3d ed. 1964). With respect to "newsworthy persons," an action for invasion of privacy must be grounded on knowing or reckless falseness in order to meet first amendment requirements. *Time, Inc. v. Hill*, 385 U.S. 374, 386-88 (1967).

⁴⁴ Cf. Bloustein, *supra* note 40.

⁴⁵ See WESTIN, *supra* note 1, at 7. See generally *id.* at 23-51.

cept where an individual's private personality to selected audience does not respect for common use of the information is proportionate or less limited and in most significant human many governmental regulations mandatory.⁴⁶

B. Threat to Privacy

A fully informed data held by federal agencies and dossiers would "k characteristics and activities of the Government,"⁴⁷ revealed in the military or to them, or been the subject

There are nearly 100 federal agencies for given disclosures relating to (1) national interests, see, e.g., INT. SEC. STAT. § 46(f) (1964) (trade secrets); (2) 80th Cong., 2d Sess. 10-11-64, has been declared to be confidential under the rule-making authority of 1670.5-8 (1968) (1670.5-8) is treated as confidential under the pledge of nondisclosure of confidential personnel and PROCEDURE OF THE SECRETARY OF THE GOVERNMENT DOSSIER 26-1-64]. See generally *id.*

varying degrees this confidential information, see 5 U.S.C.A. § 552 (1964) (McNaughton rev. ed. 1961) (552) transfer, see 44 U.S.C. § 2201 (1964) (release of data by a government agency to a private person) (44 U.S.C. § 1304(b) (1964) (al. 44 U.S.C. § 8(a) (1964) (census data) (44 U.S.C. § 6103(a) (tax information) (5 U.S.C. § 552) confidential or under regulatory control. Confidential protection is provided for information. Joint Hearings 35 (1964) (Statement of Raymond Bowman, Director of the Budget).

See generally *Hearings on the Administration of the Federal Employees Before the Senate Committee on the Judiciary*, 80th Cong.

onymity in expression⁴² b
edom of association and

a person's life, such as
clusive possession. Other
of companionship — spor
rs — with the intention
t become more widely kno
rceable by a tort action
ases it is preserved only
nfidants will respect so
disclosing the content of p
e opportunities men have
airs and the extent to wh
nication are honored may
society's regard for individu

achieved and maintained
t by existence in a crow
with others, or to withh
beliefs, opinions, affect
behavior — in the unfett
pelled exposure of this ot
esents a loss of privacy.
to a few intimates, to m
whole society, is not ne
sion of the privacy inter

ncy or gratification is subject
individuality and human dignit
His opinions, being public, t
known, tend always to be o
g openly exhibited, tend to
l to become the feelings of eve
ngible; he is not an individua
gnity: *An Answer to Dean Pr*

Investigation Comm., 372 U.S.
1960); NAACP v. Alabama ex
blatt v. United States, 360 U.S.

b) (invalidating ordinance requir

olumbia now recognize a cause
AW OF TORTS 831-32 (3d ed. 196
tion for invasion of privacy
in order to meet first amendm
, 386-88 (1967).

rally id. at 23-51.

Except where an individual has chosen to reveal an aspect of his private personality to the whole society, his disclosure to a selected audience does not forfeit a residual claim to privacy — to respect for common understandings as to what further spread of the information is proper in the circumstances. The right of more or less limited and confidential communication is implicit in most significant human relations and is a recognized necessity in many governmental requests for disclosures whether voluntary or mandatory.⁴⁶

B. Threat to Privacy Posed by a Data Center

A fully informed data center drawing on information currently held by federal agencies and capable of producing individual dossiers would "know" a great deal about the personal characteristics and activities of anyone who has ever worked for the Government,⁴⁷ received any of various special benefits, served in the military or been registered with the Selective Service System, or been the subject of any civil or criminal investigation.

⁴⁶ There are nearly 100 federal statutes which provide protection of this character for given disclosures related to personal life, see, e.g., 13 U.S.C. § 9(a) (1964) (census data); 38 U.S.C. § 3301 (1964) (Veterans' Administration records), or financial interests, see, e.g., INT. REV. CODE OF 1954, § 6103(a) (tax returns); 15 U.S.C. § 46(f) (1964) (trade secrets and names of customers). H.R. REP. NO. 1497, 89th Cong., 2d Sess. 10 (1966). Other data, not specifically restricted by statute, has been declared to be confidential and its circulation limited by regulations issued under the rule-making power of an agency. See, e.g., 32 C.F.R. §§ 1606.31-42, 1670.5-8 (1968) (Selective Service records). Finally, some information is treated as confidential by an agency because it is received under an informal pledge of nondisclosure. Some 97% of all personal disclosures are given some measure of confidential protection. See STAFF OF SUBCOMM. ON ADMIN. PRACTICE AND PROCEDURE OF THE SENATE COMM. ON THE JUDICIARY, 90th CONG., 1st Sess., GOVERNMENT DOSSIER 26-29 (Comm. Print 1967) [hereinafter cited as GOV'T DOSSIER]. See generally *id.*

In varying degrees this confidential status may immunize the disclosure from public access, see 5 U.S.C.A. § 552(b) (1967); 8 J. WIGMORE, EVIDENCE § 2377, at 781 (McNaughton rev. ed. 1961); from judicial process, see *id.*; and from intra-agency transfer, see 44 U.S.C. § 423(b) (Supp. II, 1967). Provisions for discretionary release of data by a high-ranking official are common, however, and significantly lessen the certainty of the guarantees of limited access. See, e.g., 8 U.S.C. § 1304(b) (1964) (alien registration, release by Attorney General); 13 U.S.C. § 8(a) (1964) (census data, release by Secretary of Commerce); INT. REV. CODE OF 1954 § 6103(a) (tax return available for inspection only upon order of the President or under regulations approved by him). Nonetheless it is insisted that confidential protection is an essential ingredient of a successful statistical operation. *Joint Hearings* 35 (statement of Edgar Dunn); *House Hearings* 51 (statement of Raymond Bowman, Ass't Director for Statistical Standards, Bureau of the Budget).

⁴⁷ See generally *Hearings on Psychological Testing Procedures and the Rights of Federal Employees Before the Subcomm. on Constitutional Rights of the Senate Comm. on the Judiciary*, 89th Cong., 1st Sess. (1965). See also p. 417 *infra*.

The federal government derives considerable additional information from the census, itemized tax returns, applications for issuance of a passport and documents relating to naturalization and customs.⁴⁸ Moreover, a wealth of information is currently collected and retained by agencies of state and local governments — including public hospitals, schools and universities, welfare agencies, police departments, licensing bureaus, and courts. Finally, other types of information are now in the hands of private institutions, such as banks and credit rating agencies, accountants, stock brokers, other business enterprises, social clubs, private hospitals, schools and universities, and welfare agencies.⁴⁹

The Government might conceivably move beyond the existing sources of data to create new resources. It might maintain records of individuals' incoming mail (return address and date),⁵⁰ telephone calls (numbers dialed), and expenditures. The mail and telephone surveillance could be achieved by an expansion of present capabilities. An expenditure record could be created by making government credit cards the sole legal tender.⁵¹

The capacity of a data center thus to gather together disparate bits of information about an individual might constitute a fundamental violation of expectations dependent upon the separation of those bits of information. Individual data may take on vastly increased significance when juxtaposed with other data.⁵² Much of a personality, a style of life, a network of human associations could be revealed when the separate knowledge of some twenty data-collecting federal agencies⁵³ is combined. Moreover, a computer's memory is perfect — and since it does not forget, it may not forgive.⁵⁴ And, its capacity to store and use information is vastly greater than that of any human being.

The fabric of human social relationships, dependent upon each person's having only limited knowledge about other people,⁵⁵ could be rent by leaks from the data center. The structure of social roles depends to some extent upon each participant in a

⁴⁸ See generally GOV'T DOSSIER, *supra* note 46.

⁴⁹ See generally V. PACKARD, *supra* note 7; M. BRENTON, *supra* note 7.

⁵⁰ See Senate Hearings 75 (statement of Arthur Miller).

⁵¹ Cf. WESTIN, *supra* note 1, at 163-65.

⁵² On the other hand, information obtained for one purpose may be highly misleading if used, without verification or further refinement, for another purpose. Cf. Senate Hearings 75 (statement of Arthur Miller).

⁵³ The Kaysen proposal does not specify which agencies will send data to the center but there are about 20 agencies that engage in large-scale data collection, and it is assumed that these agencies will provide the initial input. See House Hearings 2, 209-53 (inventory of data).

⁵⁴ See House Hearings 33 (remarks of Dep. Gallagher).

⁵⁵ See generally Fried, *Privacy*, 77 YALE L.J. 475 (1968).

relationship's limiting himself understood as appropriate for deal with government, even might be undermined if other party knew vastly more such a person to know. In recent years an expansion appropriate for many social example, by the use of personnel by employers, university dating services.⁵⁶ The for many purposes, both if a data center is established within government and of its wealth of information. The Kaysen Task Force proposing merely a "statistical" privacy issue the Task need not contain dossier records. Also, it been general economic, on a sample basis on named individuals. Professor center emphasizes that center would not be. This limitation, though would seem not to let fears remain that abused by over-zealous to injure particular groups might result in release more personal data; increase reliance on affect human interactions fully exposed to be a tape or punch cards of life being regulated sooner or later eventually on a broad Note, *Anthropotelemet* (1966). KAYSEN REPORT, ANNEX at See Kaysen, *Data Banks* 38. See *id.*

Not true it turns out

relationship's limiting himself to a range of concern socially understood as appropriate for the relationship. A person's ability to deal with government, employer, colleague, friend, and relative might be undermined if he had reason to suspect that the other party knew vastly more about him than it is now customary for such a person to know. To some extent, there may have been in recent years an expansion of the range of concern regarded as appropriate for many social relationships — as indicated, for example, by the use of personality tests or other informative materials by employers, universities, penal institutions, and computer dating services.⁵⁶ The undoubted utility of such information for many purposes, both benign and manipulative, suggests that if a data center is established powerful pressures will arise both within government and from the private sector to make full use of its wealth of information.

The Kaysen Task Force sought to avoid these dangers by proposing merely a "statistical" center. In an annex to the Report on the privacy issue the Task Force declared that its proposed center need not contain dossier data such as police, military, or personnel records. Also, it said that Congress could distinguish between general economic, social, and demographic data collected mostly on a sample basis and the "sort of personal history information on named individuals" that is inappropriate for a statistical center. Professor Kaysen's exposition of the proposed center emphasizes this distinction⁵⁸ and seems to say that his center would not be equipped for individualized output.⁵⁹ Assuming this limitation, the legitimate operations of the Kaysen center would seem not to invade personal privacy.

Yet fears remain that even statistical data could be used or misused by over-zealous government officials and other persons to injure particular groups; that improved capacity for data handling might result in requests by government agencies for new and more personal data; that the development of a data center might increase reliance on computers to make decisions which vitally affect human interests and which should be made by persons fully exposed to the richness and nuances of life which escape a tape or punch card and which may be crucial in the aspects of life being regulated; that individualized output might be permitted sooner or later, perhaps on a limited basis at first, but eventually on a broad and devastating scale; and that the

use of
personality
tests
was another
sign of
times.

⁵⁶ See Note, *Anthropotelemetry: Dr. Schwitzgebel's Machine*, 80 HARV. L. REV. 403, 405 (1966).

⁵⁷ KAYSEN REPORT, ANNEX at 4.

⁵⁸ See Kaysen, *Data Banks and Dossiers*, THE PUBLIC INTEREST, Spring 1967, at 52, 58.

⁵⁹ See *id.*

very existence of a data center would affect human conduct — people would so fear the possibility of exposure of their data files that they would “act for the record,” with a consequent increase in conformity. All these fears are highly generalized and speculative. Their proper assessment will depend on largely subjective judgments about attitudes and goals of public officials and powerful private individuals, about broad social tendencies in contemporary America, and about the likely course and pace of technological progress.

Finally, constitutionally protected elements of the privacy interest would be affected by a data center. A center capable of individualized output would seem to conflict most directly with the privilege against self-incrimination. If a citizen's disclosures to the Government could become available to law enforcement agencies in a manner unrelated to the justification for the initial request, opportunities for the citizen to resist any disclosure at all, without a grant of immunity, would greatly multiply. Vigorous exercise of first amendment freedoms would be jeopardized if the data center were capable of producing lists of Negro militants, political radicals, atheists, political contributors or any other potentially unpopular group. Moreover, discussion of the scope of a constitutionally protected “zone of privacy”⁶⁰ would take on new urgency if personal disclosures to the Government would be input for a massive intelligence network.

IV. ESTABLISHING A BALANCE BETWEEN EFFICIENCY AND PRIVACY

The privacy issue was largely ignored in early proposals for a data center and the Kaysen Report contained only a surface treatment. This admittedly inadequate analysis has been explained on the ground that full protection for the privacy interest was an assumed condition precedent to any successful statistical program.⁶¹ But the balance between efficiency and privacy cannot rest on an assumption. In the enabling legislation, Congress should give very careful consideration to essential legal and technological safeguards for the privacy interest. In so doing, its initial tendency should be toward overprotection; it would seem that only after a period of experimental operation with strict controls can more sophisticated distinctions be drawn intelligently without unintended sacrifices of personal privacy.

⁶⁰ See *Griswold v. Connecticut*, 381 U.S. 479 (1965).

⁶¹ See *Joint Hearings 35* (statement of Edgar Dunn).

A. I.

The most reliable safeguard implemented at the input stage is distinguished from that as input into the data, normally associated with documentary file data of the FBI, service, and medical records. Beyond specific exclusions of types of data may be Professor Kaysen has suggested input only “large-scale social statistics.”⁶² To have any effective legislative effect, however, by defining “large-scale” for different types of samples would make it more difficult to the sample's inclusion. A sample size would be of investigatory data, as a reasonably large political visibility of for the center.⁶³

The requirement of a sample to lead to very large statistical analysis, which is to be carried out, can be carried out in the interest of keeping at the center has data samples no larger than the usual associations and To facilitate comparison of samples tested for the center where the samples are

Kaysen, *supra* note 58, and under data. A class of individuals reveals data about people for the purposes of the this requirement would remain free to collect. A simple random sample any larger group in order and a tolerated error of MARCH 33 (1963). To achieve probably be considerable

A. Input Controls

The most reliable safeguards for the privacy interest can be implemented at the input stage. Only data suitable for a "statistical" as distinguished from a "dossier" center should be authorized as input into the data center. Specific kinds of data traditionally associated with dossier systems — for example, investigatory file data of the FBI and IRS as well as military, civil service, and medical records — should be expressly excluded.

Beyond specific exclusion, general principles to determine what types of data may enter the center should be prescribed. Professor Kaysen has suggested that Congress might authorize for input only "large-scale systematic demographic, economic, and social statistics."⁶² Such a standard seems too vague to achieve any effective legislative control over input. It could be improved by defining "large-scale" as a set of minimum sample sizes for different types of inquiry. Insisting on reasonably large samples would make it more difficult to relate data taken from a sample to the sample's individual members. Moreover, the minimum sample size would reinforce the ban on the center's reception of investigatory data, and by ensuring that the sample includes a reasonably large number of persons it would increase the political visibility of collection of highly personal data destined for the center.⁶³

The requirement of a minimum sample size ought not, however, to lead to very large samples since precise and reliable statistical analysis, which it is the center's main purpose to facilitate, can be carried out with relatively small samples.⁶⁴ In the interest of keeping at a minimum the number of people about whom the center has data, input to the center should be limited to samples no larger than necessary to discourage inferred individual associations and to permit maximum statistical utility.

To facilitate comparative analysis of particular variables, the samples tested for the variables would have to be identical. Where the samples are not identical, correlations of the variables

⁶² Kaysen, *supra* note 58, at 58. Professor Kaysen also suggests specific exclusion of dossier data. A class of individuals so small or homogeneous that description of the class reveals data about particular members should also be treated as an individual for the purposes of this test of "dossier data."

⁶³ This requirement would apply only to data center input. Individual agencies would remain free to collect small sample data for their own particular needs.

⁶⁴ A simple random sample need contain only 4147 representative members from any larger group in order to achieve an analysis with a confidence limit of 99% and a tolerated error of only 2%. See C. BACKSTROM & G. HURSH, *SURVEY RESEARCH* 33 (1963). To adequately safeguard anonymity each minimum sample should probably be considerably larger than this figure.

are possible only on a global basis — that is, by finding a third factor which correlates with each two factors for which different samples have been tested, an analyst can make an indirect correlation between the two factors in which he is interested. However, global comparisons constitute the very type of imprecise and unreliable analysis which the data center is intended to obviate. If the samples already used by separate federal agencies have sufficient overlap that a common general sample can be drawn without requiring additional disclosures from persons within the general sample, the use of a common sample at the data center would not create any new burdens on the selected members; it would have the desirable effect of limiting any contact with the center to the minimum necessary for statistical utility and the preservation of anonymity. However, if use of a common sample would mean that the members would be faced with a large number of new requests for data, then the idea of a common sample seems much less appealing because of the enlarged and inequitable burden it would place on the selected members. A balancing of these interests may lead to the conclusion that a set of common samples should be used according to the specific objectives of the collecting agencies.

To ensure the anonymity of persons about whom data is stored in the center, it has been suggested that individual identification be deleted at the input stage.⁶⁵ Unfortunately, this simple expedient, though effective to preclude the accumulation of dossiers, would also seriously compromise the center's statistical usefulness. Correlation of input from different sources requires the identification of all data about a person by a uniform number, especially where subsequent input may be desired. It would be feasible, however, to scramble the identifications according to a secret programmed formula. As a reinforcing safeguard, the data could be stored in the center according to general subject matter rather than in dossier format. The scattered subject tapes would be correlated only when justified by a particular program run. Congress might require that the regulating agency for the center institute such measures as these for the protection of the data's confidentiality. The suggested storage by subject matter may not be the most efficient technique, but safeguarding of the privacy interest justifies some loss of efficiency. And, since abuse of a centralized data facility will be considerably more injurious to affected individuals than prior abuse at the level of a single agency, even a substantial departure from optimum efficiency may be justified in order to make the costs of breaking the system as nearly prohibitive as possible.

⁶⁵ See House Hearings 14, 18 (testimony of Vance Packard).

him.
Split up
data on
different
tapes to
make it
physically
difficult to
violate
privacy

the central concern
center within its pri
tical resources and
city. The major li
technological ban c
lations could distin
gh some individual
the possibilities a
tion to the general
prohibit all individu
ment of the ce
limitation the s
ating output wit
ls, but also incap
population is less t
be to prevent
tion so small t
easily be inferre
useful and legiti
could be used b
ualized output v
mental data.

further safeguar
control. In addi
ade for the mai
of the use of
67

C. A

data center wh
ive of that hel
ble of individu
to personal pri
tion of privacy
es where they
abling legislatio
to provide ad

House Hearings 9
COMPTROLLER GE
No. 224, 90th C
United States Compt
ANNUAL REPORT].
753, tit. IV, 60 S
(C.).

B. Output Controls

The central concern in safeguarding privacy should be to keep the center within its primary role of improving the Government's statistical resources and to prevent the development of a dossier capacity. The major limitation on the center should be a legal and technological ban on any individualized output. Statutes or regulations could distinguish lawful from unlawful output. Even though some individualized output might serve a legitimate purpose, the possibilities and costs of abuse of even a well-defined exception to the general rule are so great that it seems preferable to prohibit all individualized output, at least during the early development of the center. To assure the effectiveness of the output limitation the system could be made not only incapable of printing output with individual names or other identifying symbols, but also incapable of printing output in any case where the population is less than a programmed minimum.⁶⁶ The aim would be to prevent any output which might be based on a population so small that information about given individuals could easily be inferred. To the extent that individualized output is useful and legitimate for particular purposes, smaller computers could be used by individual agencies which when making individualized output would not have access to the whole store of governmental data.

A further safeguard against improper output is effective security control. In addition to physical security, provision should be made for the maintenance of detailed records and random audits of the use of the center to deter possible abuse from within.⁶⁷

C. Additional Legal Safeguards

A data center which possesses only large-scale data representative of that held by individual agencies and is physically incapable of individualized output would seem to pose little danger to personal privacy. Since such controls cannot guarantee protection of privacy, however, a remedy should be provided in cases where they fail. Congress should consider as part of the enabling legislation an amendment to the Federal Tort Claims Act⁶⁸ to provide adequate compensatory relief for individuals

⁶⁶ See *House Hearings 94* (testimony of Edgar Dunn).

⁶⁷ Cf. *COMPTROLLER GENERAL OF THE UNITED STATES, ANNUAL REPORT 1967*, H.R. DOC. NO. 224, 90th Cong., 2d Sess. (1967) (audit procedures and operations of the United States Comptroller General) [hereinafter cited as *COMPTROLLER GENERAL ANNUAL REPORT*].

⁶⁸ Ch. 753, tit. IV, 60 Stat. 842 (codified, as amended, in scattered sections of 28 U.S.C.).

first look
in Dejeune's
was dossier
King Mads
CD 200
EPA/FOX

IRS
Bowsing

instead, we
got
FOIA

whose privacy has been invaded by means of the center and the safeguards. The *in terrorem* effect of civil liability could be enhanced by conferring upon members of the public advisory council standing to institute appropriate administrative or judicial actions.

Several commentators have urged that some provisions be made to assure that possible abuse of the center will be based at least on accurate data. To this end it has been suggested that individuals about whom the center has information be given annually an opportunity to review such information.⁶⁹ But the provision for individual review requires a capacity for individual output, the absence of which is one of the major safeguards of the privacy interest. Moreover, the Government's interest in statistical accuracy seems greatly outweighed by the potential threat to individual anonymity and by the costs of instituting any effective procedure. Finally, an individual's interest in privacy is significant only where the data is intended for individual use; where such administrative data is stored in separate level computers, review by interested individuals could be done there without jeopardizing a fundamental limitation on the central data facility.

V. CONCLUSION

Whatever the precise solutions that prove most workable, it is incumbent on Congress to anchor general safeguards of accuracy and confidentiality in the data center's enabling legislation and to charge the facility's regulating agency with the duty to propound and enforce more particular measures. A public advisory council such as that recommended by the Kayser Commission but responsible directly to Congress⁷⁰ could be an effective representative of the citizen's interest in privacy. A data center properly so limited could add substantially to the effectiveness of governmental statistical operations and offer fruitful data for research without any significant inroad into personal privacy.

Much of the criticism of the data center proposal seems to be in part a product of a growing concern that expansive government and the intrusive requests for data accompanying its growth are seriously limiting the opportunities for privacy of individuals.

⁶⁹ See, e.g., *House Hearings* 3-4 (statement of Rep. Gallagher); *Senate Hearings* 77 (statement of Arthur Miller); cf. Karst, "The Files": Legal Control of the Accuracy and Accessibility of Stored Personal Data, 31 *LAW & CONTEMPORARY PROBLEMS* 342, 358-59 (1966).

⁷⁰ Direct responsibility exists in the case of the General Accounting Office, which in many respects performs analogous functions. See generally *CONGRESSIONAL GENERAL ANNUAL REPORT*, *supra* note 67.

1
Don't let people
review their
data

When Co
fully reev
There is
citizens at
tion. Per
should no
Senate ha
would ma
disclose p
submit to
stances.⁷¹

Althou
center see
ernment is
some, the
fears with
the sacrific
developme
which pote

⁷¹ S. 1035,
90th Cong., 1

means of the center des...
ect of civil liability could...
bers of the public advi...
riate administrative or o...
ged that some provision...
of the center will be base...
d it has been suggested...
r has information be...
such information.⁶⁹ But...
a capacity for individual...
e of the major safeguard...
he Government's interest...
outweighed by the pote...
l by the costs of institu...
individual's interest in ac...
ta is intended for indivi...
is stored in separate ag...
ed individuals could be...
mental limitation on the

When Congress considers the data center proposal it might usefully reevaluate the entire scope of governmental data collection. There is an interest in privacy that should be weighed whenever citizens are required by statute or regulation to provide information. Personal data not essential to governmental operations should not be collected except on a strictly voluntary basis. The Senate has recently recognized this policy by passing a bill that would make it unlawful to require a Government employee to disclose private activities unrelated to his official duties or to submit to psychological tests except under very limited circumstances.⁷¹

Although public concern may delay its creation, a federal data center seems an inevitable response to the needs of effective government in an increasingly complex society. Because it is awesome, the computer is feared. The proper response to present fears with respect to a computerized data center, however, is not the sacrifice of the computer's positive contributions, but rather development of the techniques, both legal and technological, by which potential abuse may be prevented.

⁷¹ S. 1035, 90th Cong., 1st Sess. §§ 1(d)-(e) (1967). See also S. REP. No. 534, 90th Cong., 1st Sess. (1967).

SION

hat prove most workab...
- general safeguards of...
center's enabling legis...
ng agency with the du...
lar measures. A publi...
ended by the Kaysen...
gress⁷⁰ could be an effe...
st in privacy. A data c...
tentially to the effectiv...
s and offer fruitful dat...
road into personal pri...
ta center proposal seem...
that expansive governm...
accompanying its growth...
for privacy of individual

of Rep. Gallagher); *Senate Ho...*
1, "The Files": Legal Controls O...
nal Data, 31 LAW & CONTEMP. P...

of the General Accounting O...
ctions. See generally COMPTE...