

A STUDY OF THE MOVEMENT FOR
A NATIONAL STATISTICAL DATA CENTER (1960-1967)

by

WALTER YORSZ

S.B., Massachusetts Institute of Technology
(1970)

SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF

SCIENCE

at the

MASSACHUSETTS INSTITUTE OF

TECHNOLOGY

May, 1976

Signature of Author.....
Department of Political Science, May 7, 1976

Certified by.....
Thesis Supervisor

Accepted by.....
Chairman, Departmental Committee on Graduate Students



A Study of the Movement for
A National Statistical Data
Center (1960-1967)

by

Walter Yorsz

Submitted to the Department of Political Science on
May 7, 1976, in partial fulfillment of the requirements
for the degree of Master of Science.

ABSTRACT

In the early 1960's a group of economists proposed to the Bureau of the Budget the creation of a computerized national statistical data center as an attempt at improving the efficiency of federal statistical operations.

After an examination by the Bureau of the Budget and with the publication of its own report on the data center concept, the issue attracted widespread public concern with regard to its possible violation of personal privacy rights. A series of Congressional investigations and the ensuing public criticism of the concept successfully stopped its implementation.

This paper examines the history of this issue with special attention to the differing perspectives of the executive and the legislative branches of the federal government in examining and evaluating the proposal for a national data center.

Thesis Supervisor: Ted Greenwood
Assistant Professor, Dept. of Political
Science

ACKNOWLEDGMENTS

I would like to express my thanks to the many individuals who were interviewed in the course of this study for their invaluable information. I would also like to thank Ted Greenwood for his continual guidance and support without which this study would not have been possible. Special appreciation goes to my wife whose skills as an editor were put to the supreme test with this document.

TABLE OF CONTENTS

	Page
ABSTRACT	2
ACKNOWLEDGEMENTS	3
LIST OF TABLES	5
PREFACE	6
CHAPTER I. INTRODUCTION.....	7
CHAPTER II. HISTORICAL REVIEW.....	9
CHAPTER III. ISSUES WITHIN THE EXECUTIVE BRANCH.....	22
Definitions.....	22
The Federal Statistical System and The Movement for Reform.....	25
The Impact of Computer Technology.....	32
The National Data Center.....	40
Another Perspective on the Data Center.....	43
The Kaysen Report and Its Implications.....	50
Prior to the Congressional Hearings.....	54
CHAPTER IV. THE GALLAGHER HEARINGS.....	56
The Privacy Issue and the Gallagher Committee.....	56
The Witnesses and the Issues.....	58
Responses From the Executive Branch.....	63
Comments on the Hearings.....	68
CHAPTER V. CONCLUSIONS.....	71
Aftermath of the Gallagher Hearings.....	71
Conclusion.....	72
NOTES	77

LIST OF TABLES

<u>Table</u>	<u>Page</u>
1. Obligations for Principal Statistical Programs Selected Years, 1934-1971	29

PREFACE

Unless otherwise referenced, the information contained in this paper was derived from a series of interviews with the following individuals who were involved in the data center issue: Richard Ruggles, Edgar Dunn, Guy Orcutt, Ed Goldfield, Richard Tayber, Morris Hansen, Charles Zwich, and Charles Schultze.

CHAPTER I: INTRODUCTION

Among the more impressive accomplishments of the past few decades has been the remarkable evolution of technology related to the storage, processing, and transmittal of information. The transition from manual to computerized information systems has resulted in an expansion of data processing capabilities whose dimensions are without parallel in the history of technological development. The rapidity and scope of the transition far outstrips the shift from chemical to nuclear energy, and in a society where information is, in John McCarthy's words, "a commodity no less tangible than energy, and even more pervasive in human affairs,"¹ the consequences are likely to be equally significant.

We are currently witnessing the beginning of a process in which all types of records, public and private, are being converted to computer readable format, and it is likely that within the next few decades virtually all of the world's recorded information will be available in digital form. While the efficiency and economic advantages of this trend are enormous, its impact on human liberties and desires is not yet clearly perceived.

One of the earliest important debates regarding this potential conflict took place in the early 1960's over the proposals for the establishment of a national statistical data center. This paper

examines the evolution of these proposals, their evolution and subsequent rejection, within the executive and legislative branches of the federal government. The objective was to determine how one of the first major repudiations by the federal government of technological advance came about.

CHAPTER II: HISTORICAL REVIEW

The first step toward the formulation of proposals for the creation of a national data center came as the second part of a dual proposal submitted to the American Economic Association by the Ford Foundation. In its proposal the Foundation offered to fund two activities: the first was the establishment of a journal of abstracts of current economic literature; the second was the formation of a user oriented group which would act to communicate the data and data base requirements of the economics community to the federal government.

The proposals evolved from a concern on the part of the Foundation with the progress of economic research, and specifically from the work of Dr. Richard Ruggles, a consultant to the Foundation and head of the Economics Department of Yale University, whose principal responsibility at Ford was the identification of problem areas relating to such research. Ruggles defined two areas which, at the time, represented major concerns to economists: the first was the communications problem arising from the proliferation of economic literature and periodicals; the second was the growing dissatisfaction of economists with the quality and availability of basic statistical data compiled by the federal government -- hence the proposals to the AEA.

The AEA accepted the first part of the Ford offer and the Journal of Economic Abstracts (now the Journal of Economic Literature) was created. With regard to the second part, the AEA's own organizational structure prevented its addressing in any substantive fashion the issues involved. It therefore sought the assistance of the Social Science

Research Council in establishing a committee to study the problems associated with the use of economic data and to undertake any program of action which might seem appropriate.²

Accordingly, in December 1960, Henry Regan, then president of the SSRC, arranged an exploratory discussion of the problems of data use in the social sciences, drawing on social scientists from the universities and the federal government. This discussion revealed that although there were large areas of common interest among the various social sciences, there were also wide differences in approach and concern when specific aspects of data use were considered. As a consequence, the scope of the discussions was narrowed to the initial concern of the AEA and the Ford Foundation with the problems of economic data.

To investigate these in greater detail Dr. Ruggles was asked to appoint and chair a special committee on the preservation and use of economic data. He chose four other academic economists to complete the committee.³ As the focus of the committee's efforts was to be on data collected and maintained by the federal government, all five individuals were well acquainted with the work of the federal statistical system and were also involved with research activities requiring large bodies of statistical data.

At the outset the committee chose to focus on areas which could yield the most valuable research material per unit cost and thus directed its attention to those bodies of data which were available in machine readable format. The two principal concerns of the committee were the dissemination and preservation policies of the federal government with respect to this data.

During 1962 to 1965, the committee undertook to study on an agency-by-agency basis the problems relating to the preservation of, and access to, specific bodies of information. Meetings were held with a considerable number of agencies in the Departments of Commerce, Labor, Treasury, Agriculture, Interior, and Health, Education, and Welfare. In addition the committee kept close contact with the Bureau of the Budget, the National Archives, and the National Science Foundation. In April of 1965, the committee submitted its final report to the Social Science Research Council.⁴ In essence it concluded that the growing decentralization of the federal statistical system was increasingly hampering the effective utilization of government statistics.⁵ In particular it cited the following major problems as arising from the existing decentralization system:⁶

Preservation of Data. Under a decentralized system it is extremely difficult to maintain a coherent and consistent policy with respect to the preservation of basic data. Where various agencies are primarily responsible for day-to-day operations, they frequently do not, or cannot, give high priority to long-run considerations. Thus, there is no adequate mechanism for insuring that these agencies are following optimal policies with respect to the preservation of important data.

Data Access. Owing to the decentralized nature of the federal statistical system, it is extremely difficult for potential data users outside the federal system, or even other federal agencies, to discover precisely what data tapes exist and the procedures for

gaining access to them. Moreover, as the production of data is an activity ancillary to the primary responsibilities of most federal agencies, they find it both inconvenient and costly to respond to requests for specific information which might disrupt their normal operations.

Development of Usable Data. Under the present decentralized system emphasis is placed on the development of specific tabulations or computations relevant to the operations of the agency producing the data. Thus supporting and explanatory data relevant to these final tabulations, which might make the data useful to other agencies, is often not prepared or made available in a clean edited form. The value of the data is thus restricted to the producing agency.

Disclosures. In addition to the problems of locating and gaining physical access to the data, a considerable portion of the data obtained by the federal government from individuals and firms is done so with the understanding that it will be considered confidential and not available to those outside the government or even other federal agencies. While all concerned have recognized that no federal agency ought to violate this confidentiality, agencies often do provide outsiders with statistical summaries of specific data in such a way that individuals cannot be identified. However, with the exception of the IRS and the Census Bureau, few federal agencies have the manpower or funds to provide users with such services. Moreover, restrictions on interagency disclosures preclude the performance of many interesting correla-

tions of data prior to the publication of statistical summaries.

In response to these perceived deficiencies the Ruggles committee suggested greater centralization on the part of both the users' community and the federal statistical system. Specifically, the committee offered three recommendations:⁷

- 1) The Bureau of the Budget, in view of its responsibility for the federal statistical program, should immediately take steps to establish a National Data Center.
- 2) The Office of Statistical Standards of the Bureau of the Budget should institute programs to insure the systematic preservation in usable form of important data by those agencies engaging in statistical programs.
- 3) At an early date the Social Science Research Council should convene representatives from research institutions and universities in order to develop an organization which can provide a clearing house and coordination of requests for data made by scholars from federal agencies.

The data center suggested by the Ruggles group would have as its primary objective the provision of basic data regarding the American economy to individual users or other federal agencies in whatever format they desired within the bounds of a single, coherent disclosure policy. It would also provide, on a reimbursable basis, production runs and aggregated tapes and results to scholars, thereby eliminating inter-agency disclosure problems. It would thus serve a role similar to that

of the Library of Congress providing systematic and comprehensive coverage of data in its area of competency. In order to carry out its function it would necessarily have both substantial computer facilities of its own and interagency authority to obtain basic data in machine readable format from any federal agency engaged in statistical operations. The Center could thus collect whatever data might be relevant to its operations and:

- catalog and make known to the user community the full range of available data,
- develop consistent policies regarding the preservation of important data,
- enforce uniform access and disclosure procedures, and
- make the fullest use of available data with carefully designed file maintenance and documentation procedures.

The Ruggles committee urged the Bureau of the Budget to take the initiative in promoting the establishment of the National Data Center. Charles Schultze, director of BoB at the time, proved receptive to the committee's suggestions, and at the urging of his office, Raymond Bowman, then head of the office of Statistical Standards, commissioned a second study of the issues raised by the Ruggles report. To conduct the study, Bowman chose Dr. Edgar Dunn, a consultant with Resources for the Future and formerly Deputy Assistant Secretary of Commerce for Economic Affairs, known within the federal statistical community for his initiation of important statistical reforms within the Census Department.⁸

The Dunn report was completed in approximately six months of part-

time effort by Dr. Dunn and an informal five-man advisory committee and was submitted to the Bureau of the Budget in November of 1965.⁹ This report concurred fully with the Ruggles committee's assessments of the problems of the federal statistical system, noting also in passing that concern with the statistical system extended beyond the academic social science research community, encompassing among others numerous federal agencies such as the Office of Economic Opportunity, the Department of Health, Education, and Welfare, the Department of Urban Affairs, etc., all experiencing great difficulties in assembling the statistical data to guide them in analyzing their missions, establishing performance standards, and supporting administrative decisions and the evaluation of results.

Beyond these arguments for the necessity of some form of action, Dunn focused on specific considerations relating to the implementation of a National Data Center. He began by taking exception to the term "National Data Center" on the grounds that it suggested an overly superficial approach to the statistical problem and proposed instead the creation of a National Data Service Center, choosing the term carefully to emphasize the service and coordinating functions of the proposed facility and to differentiate his concept from that of a simple computerized data file.¹⁰ Dunn saw neither a pressing requirement for nor any advantage deriving from the centralization of federal statistical files (with the exception of those with archival value which might not ordinarily be preserved by the agencies which originally developed them). Instead, he was the solution to the problems of the statistical system in terms of the promulgation throughout the system of uniform standards of file

design and maintenance in order to insure maximum file capability.

Specifically, Dunn felt the major deficiency of the federal statistical system was the difficulty in associating records obtained from different statistical files, thus precluding the identification and examination of many interesting interrelationships. At the heart of the problem was the incompatibility of file structures, i.e. different formatting or classification schemes used by different agencies. The physical centralization of incompatible files would serve little useful purpose, and the crucial requirement was a mechanism for the assurance of interagency file compatibility. Having achieved such compatibility, the compactness and portability of magnetic tape data storage would eliminate the need for physical centralization of any but the most frequently used data files.¹¹

However, this important distinction between the Ruggles and Dunn perspectives later tended to be obscured, in part because both reports considered a broader range of issues than just file centralization but primarily because Dunn also recommended the creation of a data center, albeit one of limited scope intended to hold only those files which either were of considerable archival value or which were likely to be used very frequently in the course of performing user requested analysis. To this end, the report contained in one of its appendices a listing of 9,000 computer tapes which might be used to form the core of such a center.¹² While Dunn later explained that the centralization of these tapes was proposed only so that they might be preserved until it could be determined which were of sufficient value to include in an operational center,¹³ this point was not made sufficiently clear in the initial

report, and as a consequence the 9,000 tapes were interpreted as only the first step in the compilation of a much more massive data center.¹⁴ Dunn thus came to be identified with Ruggles as an advocate of a massive National Data Center.

Nonetheless, the distinction between Dunn's view and his own was not lost on Ruggles who was committed to the large scale centralization of the federal statistical files. He therefore sought support for his position and was able to persuade Charles Schutze to appoint a third commission, under the direction of Carl Kaysen, a Harvard economist known by Ruggles to be sympathetic to his views, to further investigate the statistical system.¹⁵ Kaysen subsequently appointed five economists -- including Dr. Ruggles -- to form the Task Force on the Storage of and Access to Government Statistics.¹⁶

While the task force went about its work, Congressional interest developed in the data center question. The issue was first brought before Congress with the testimony of Edgar Dunn before the Senate Subcommittee on the Judiciary, chaired by Senator Long of Missouri.¹⁷ The overall topic of the subcommittee's investigation was the invasion of privacy, with interest being principally focused on the increasing use of electronic surveillance devices, especially by the government on its own employees. While the data bank may have been relevant to its proceedings, the committee's interest was, at the time, peripheral as it called only a single witness (Dr. Dunn) to testify before it. Nonetheless, the subcommittee's interest did serve to spread discussion of the issue outside the bounds of the federal statistical community and led to further discussion of the proposals both within Congress and by

the public.

The Senate Subcommittee's hearings were held in June of 1966; a month later the data center was considered in greater detail during two days of hearings by a Subcommittee of the House Committee on Government Operations, chaired by Cornelius Gallagher of New Jersey.¹⁸

Gallagher had, along with Senators Long and Ervin, a reputation as being among the staunchest advocates of personal privacy rights within the Congress.¹⁹ Having recently concluded a series of investigations regarding the use of personality tests by federal agencies,²⁰ he turned his attention to other trends within the government which he perceived as constituting possible threats to personal privacy, coming to focus on the data bank proposal. The Congressman expressed concern that the Budget Bureau in its desire to improve the efficiency of government operations had failed to give adequate consideration to the implications of such a proposal in terms of potential privacy intrusions. What he sought was the creation of "a climate of concern, in the hope that guidelines can be set up which will protect the confidentiality of reports and prevent the invasion of privacy, while at the same time allowing government to function more efficiently and facilitating the necessary research of scholars in statistical analysis."²¹

Despite Gallagher's expressed concern for the necessity of a balanced perspective, his own view, and the tone the hearings would subsequently take, was evidenced in remarks such as the following:²²

The presence of these records in government files is frightening enough, but the thought of them neatly bundled together into one compact package is appalling. We cannot be certain that such dos-

siers would always be used for benevolent purposes....The possible future storage and regrouping of such personal information also strikes at the core of our Judeo-Christian concept of "forgive and forget," because the computer neither forgives nor forgets...such power would enable a less scrupulous person -- or even a well meaning but overzealous government official -- to delve behind the statistic to the respondent and learn the inner secrets of the individual.

A series of witnesses from the academic world, the government, and the legal profession, called before the committee to offer their views, tended on the whole (with the exception of witnesses for the Bureau of the Budget) to reflect Gallagher's perspective, expressing serious reservations regarding the successful implementation of the proposed data center without substantial intrusions on personal privacy.²³

For its part, the Budget Bureau had largely failed to anticipate the extent of the emotional reaction that the proposition evoked.²⁴ The privacy issue which came to be the focus of Congressional and public concern had been virtually ignored in the reports and discussions within the executive branch. This is not to say that the proponents of the data center were unaware of the privacy implications, but the implicit assumption had been that privacy, which had been very well protected by the administrative procedures of executive organizations such as the Census Bureau, could be protected equally well in a data center operating under the same procedures.²⁵ However, both the Gallagher Committee and later the public press viewed the omission of explicit discussions of privacy in the Ruggles and Dunn reports as indications of a lack of serious concern for the issue within the executive branch.

Thus while the BoB reports, and to a large extent, its testimony

before Congress tried to focus on improving the efficiency of the federal statistical system, Gallagher and the press directed their concern to the possible Orwellian implications of the proposal. Gallagher saw the threat of a computerized man, stripped of his individuality; the press saw in the center the beginnings of a police state -- a computerized Gestapo.²⁶

The impact of the near unanimous press condemnation coupled with Congressional pressure was sufficient to halt any further consideration of the program. The concept itself was, at the time of the hearings, in a sufficiently embryonic stage that it had not attracted either the attention or the support of any but a few members of the executive branch whose commitment was half-hearted at best.²⁷ Thus for all practical purposes the concept died with the Gallagher hearings and the ensuing publicity in the autumn of 1966. The submission of the Kaysen report in October of that year, again advocating a data center, revived the issue briefly but a number of factors contributed to prevent any renewed push.²⁸ Magazine articles and editorials in the press remained critical of any attempt to centralize government data, and Congress maintained its own pressure on the Bureau of the Budget with two sets of hearings before Senator Long's Subcommittee on Administrative Practice and Procedure in March of 1967 and February of 1968.²⁹ With the worsening of the Vietnam War in late 1967 and the general demoralization of the federal bureaucracy that followed little enthusiasm could be generated within the Bureau of the Budget for marshalling the time or resources necessary to adequately respond to Congressional pressure on the data bank. The election of a new administration in 1968, with a clear

philosophy against government centralization, effectively shelved any further consideration of the issue.

CHAPTER III: ISSUES WITHIN THE EXECUTIVE BRANCH

Definitions

The movement for a National Data Center was the product of a growing concern on the part of a certain class of data users with the efficacy of the federal statistical system. It is worthwhile at the outset, therefore, to distinguish between statistical and other types of data gathered on individuals since the recognition of these distinctions had a bearing on the perspectives that various groups brought to the examination of the data center issue. Westin defines three types of records pertaining to individuals:³⁰

Administrative Records. The administrative record is often generated in the process of a transaction -- marriage, graduation, obtaining a license or permit, buying on credit, or investing money. Usually a record that refers to an individual includes an address or other data sufficient for identification. Personal data in an administrative record tends to be self-reported or gathered through open inspection of the subject's affairs. Private firms usually treat administrative records pertaining to individuals as proprietary information, while administrative records held by the government are normally accessible to the public and may be shared for administrative purposes among various agencies. Administrative records sometimes serve as credentials for an individual; birth certificates, naturalization

papers, bank records, and diplomas all serve to define a person's status.

Intelligence Records. The intelligence record may take a variety of forms. Familiar examples are the security clearance file, the police investigative file, and the consumer credit report. Some of the information in an intelligence record may be drawn from administrative records, but much of it is the testimony of informants and the observations of investigators. Intelligence records tend to circulate among intelligence-gathering organizations and to be shared selectively with organizations that make administrative determinations about individuals, such as the granting of employment or security clearance. Intelligence records are seldom deliberately made public, except as evidence in legal proceedings. The largest categories of intelligence files are those maintained by law enforcement agencies, principally on the federal and local levels.

Statistical Records. A statistical record is typically created in a population census or sample survey. The data is usually gathered through a questionnaire or by some other method designed to assure the comparability of individual responses. In nearly all cases, the identity of the record subject is eventually separated from the data in the record. If a survey must follow a given individual for a long time, his identity is often encoded, with the key to the code entrusted to a separate record to guard anonymity. Data from administrative records are sometimes used for statistical purposes, but statistical records about identifi-

able individuals are not used for administrative or intelligence purposes. Thus Internal Revenue Service records and Census Bureau records contain information on individuals; however, only statistical summaries of such information -- in which individuals cannot be identified -- are ever released.

Not every type of record or data can be fit unambiguously into one of these three categories. Statistical records are sometimes derived from administrative files and do at times contain sensitive information. Nonetheless, these are viable distinctions. Those engaged in the production and maintenance of data files have, in general, little difficulty in categorizing the data they handle as either administrative, intelligence, or statistical. Moreover, tradition dictates that the three types of records should be held separately, and each used for its nominal purpose only.³¹ The transfer of data from one type of record to another takes place only under controlled conditions.

These procedures and distinctions were implicit in the thinking of the proponents of the data center. Neither they, nor other members of the federal statistical community ever envisaged the mass transfer of data from administrative and intelligence files into a national statistical data center. This was, however, precisely the image formed by many outside the statistical community who were unfamiliar with its traditions, and the failure of the center's proponents to sufficiently delineate before Congress and the public the scope of these traditions contributed enormously to the criticisms the data center proposals encountered.³²

The Federal Statistical System and the Movements for Reform

The movement for a national data center was only the latest in a long series of commissions and studies, dating from 1844, directed at improving the efficacy of the federal statistical system, and many of the concerns and recommendations voiced by Ruggles, Dunn, and Kaysen represented more or less perennial comments.³³

Inefficiency stemming from decentralization and subsequent duplication of effort has always been a popular complaint, and increased centralization is usually the proposed solution. The first movement toward centralization and coordination of effort took place with the creation by Congress of a Bureau of Statistics within the Treasury Department in 1866. In part this action was the result of the first Congressional study of the federal statistical system commissioned by the House of Representatives in 1844.

In 1903, the transfer of the Bureau of Statistics to the newly created Department of Commerce and Labor resulted in the immediate appointment by the Secretary of a special commission to consider the reorganization of statistical work within the department. The recommendations of the commission calling for increased centralization of statistical activities were not carried out, nor were the similar recommendations that were presented to the Secretary's successor in 1908 by another special commission.

By 1918, the war effort had resulted in a new emphasis on efficiency and centralization. The Central Bureau of Planning and Statistics had been established to review and consolidate forms, build up an

index of all data collections by the government and promote standardized classifications. Its functions were subsequently taken over by the Bureau of Efficiency in 1919, when that agency was asked by Congress to undertake a thorough review of the federal statistical system. Completed in 1922, the review recommended a reorganization of statistical operations to concentrate, as far as possible, collection, compilation, and dissemination of all nonadministrative statistics in a central bureau because under the present system

- 1) user access was difficult,
- 2) uncoordinated surveys were burdensome to respondents, and
- 3) cost was excessive.

No action was taken on the report.

In 1933, the funding of the government statistical program was jeopardized by reductions in federal expenditures. As a consequence the American Statistical Association and the Social Science Research Council combined to form the Committee on Government Statistics and Information (COGSIS) to advise government on organizational matters with respect to the statistical system. Although not established until mid-June of 1933, it was able to recommend the creation of a Central Statistical Board by July 1933, and see it established within a matter of weeks by executive order. COGSIS urged the expansion of the Central Statistical Board as a coordinating agency, but with only advisory power and not the authority to command agencies to act. In 1940, the Central Statistical Board was converted into the Division of Statistical Standards and placed within the Budget Bureau (the con-

version involved no major change in function or powers). The powers of the Statistical Division were expanded somewhat during the war effort when paperwork burdens being placed on manufacturers resulted in the Federal Reports Act of 1942 which gave the Division four powers:

- 1) To forbid the use of overly complex forms,
- 2) To designate a specific agency to perform specific work,
- 3) To forbid agencies to collect unnecessary information (the definition of unnecessary was left to BoB), and
- 4) To force agencies to divulge information they had gathered to other agencies.

In 1948, as part of the overall work of the first Hoover Commission, the National Bureau of Economic Research was asked to produce a report on the statistical system of the federal government. The report, authored by F. C. Mills and Clarence Long, noted the following as some of the problems of the existing systems:

- Overlapping jurisdiction of agencies
- Lack of balance in resources allocation
- Lack of comparability of data from different sources
- Variations in quality
- Poor means of access by users
- Gaps in coverage of some areas
- Methodological blunders
- Publication time and adequacy shortcomings

As with previous reports greater centralization was urged as a means

for overcoming these difficulties. Mills and Long also recommended an expansion of the powers of the Division of Statistical Standards to enable it to better carry out its coordinating responsibilities and increase the overall efficiency of the statistical system. As a consequence, Congress, in 1950, did further expand the authority of the Division of Statistical Standards to develop programs and issue orders for improving the statistical activities of the executive branch. But immediately following the Korean War, the staff of the Division was cut from sixty to fewer than forty. With the reduced manpower level, the Division was never able to pursue its coordinating activities very vigorously.

However, with respect to the centralization of operational activities, the Mills-Long recommendation, like the 100 years of similar recommendations which preceded it, had little impact. In fact, the trend in the federal statistical system has been, especially in recent years, one of growing decentralization. In 1934, the three major statistical agencies within the federal government, the Census Bureau, the Bureau of Labor Statistics, and the Department of Agriculture (which contains the Economic Research Service and the Statistical Reporting Service), accounted for 92% of federal expenditures on major statistical programs. At the writing of the Mills-Long report, their share was 78%, and at the writing of the Kaysen report in 1966, it would be 60%.³⁴ (The table on the following page lists the expenditures for selected years of some of the major federal statistical agencies.³⁵)

The trend toward a decentralized statistical system in spite of continued recommendations for increased centralization is the product

TABLE 1 OBLIGATIONS FOR PRINCIPAL STATISTICAL PROGRAMS,
SELECTED YEARS, 1934-1971

<u>Title of Agency or Series</u>	Fiscal Years (millions of dollars)				
	<u>1934</u>	<u>1939</u>	<u>1948</u>	<u>1961</u>	<u>1971(est)</u>
Total obligations -- all					
principal stat. programs	10.6	15.7	30.7	67.2	221.3
Periodic Census -type pro-					
grams-annual coverage	5.1	5.1	7.5	18.8	30.0
Current programs-total	5.5	10.6	23.2	48.4	188.3
Bureau of the Census	2.4	2.1	7.4	9.6	22.6
Statistical Reporting					
Service and part Economic					
Research Service (Agric.)	1.7	3.8	4.9	9.8	23.4
Bureau of Labor Statistics	.6	2.1	4.3	11.1	26.4
National Center for Edu-					
cational Statistics	-	-	.1	.9	5.7
National Center for Health					
Statistics	-	-	1.0	4.0	10.1
Bureau of Mines	.2	.3	.7	2.0	3.6
Office of Business Economics	-	.3	1.1	1.5	4.0
Social Security Admin-					
istration	-	.8	1.2	2.6	13.4
Social and Rehabilita-					
tion Service	-	-	-	-	4.9
Department of Housing					
and Urban Development	-	-	-	-	3.0
Department of Transporta-					
tion	-	-	-	-	5.7
Internal Revenue Service	.3	.5	1.1	3.1	8.0
Interstate Commerce Com-					
mission	.2	.6	.7	1.2	.7
Office of Equal Oppor-					
tunity	-	-	-	-	4.9

TABLE 1 (cont.)

<u>Title of Agency or Series</u>	Fiscal Years (millions of dollars)				
	<u>1934</u>	<u>1939</u>	<u>1948</u>	<u>1961</u>	<u>1971(est)</u>
Other agencies	-	-	.7	2.6	51.9
Other associated statisti- cal series: Statistical Policy Division	NA	.125	.373	.457	2.711

Totals may not add because of rounding.

of a number of factors. First of all, the commissions and study groups examining the statistical system tended to suboptimize. That is, they focused less on the difficult questions of what data the government ought to be obtaining and supplying users, than on how existing services might be expanded with only marginal cost increase. This approach overlooks the fact that the implementation and administration of new government programs frequently requires the compilation of whole new classes of data which are unavailable on existing records. As a consequence a new data gathering activity is established and is usually incorporated within the structure of the agency charged with the overall administration of the program in question. The proliferation of government social programs -- and the subsequent requirement of statistical data for the administration and evaluation of those programs -- thus significantly contributed to the decentralization of the statistical system. Subsequent attempts to divorce these statistical activities from their parent agencies and bring them under a single centralized operation encountered the charge that in so doing they would be less responsive to the original parent group and thus fail to perform with optimal efficiency the task for which they were created.

Beyond this, charges of inefficiency in the federal statistical system tended to be discounted by the members of that community. While potential improvements could always be identified, the statistical community could justifiably feel that U.S. government statistics were of the highest quality to be found anywhere in the world. Innovations in methodology and processing were adopted without unreasonable

lags and occasionally were developed by the statistical services themselves. Moreover, the calls for centralization were seen by many as less the result of identifiable system failure than the natural proclivity of the human intellect toward "orderliness". That is, a decentralized system is seen as "disorderly" to the casual observer while a centralized system is orderly and efficient.

Given this background, it is not surprising that the reports of Ruggles, Dunn, and later Kaysen, generally evoked little response within the statistical system. To the extent that they expressed concerns regarding the inefficiencies of decentralization they merely echoed the sentiments of numerous previous reports which had, in the final analysis, little impact on the organization of the system. In the respect that the reports were unique -- their recommendations for some form of computerized data center -- they were not sufficiently relevant (at least in the case of Ruggles and Dunn) to be of interest to most of the agencies. That is, neither Ruggles nor Dunn specifically discussed the impact of a data center on the organizational structure of the statistical system. The concept of a data center had been informally discussed within the statistical community for a number of years, and in large measure, these reports were seen as an extension of those discussions. Until the data center proposals became sufficiently well defined that they began to encompass organizational considerations most of the statistical agencies were far too occupied with other matters to take an active role in the consideration of the issues.

Parenthetically, the Kaysen report did address organizational

considerations, but at the time of its appearance in October of 1966, the data center was already a dead issue.³⁶ The report thus received little serious consideration from the statistical community.

The Impact of Computer Technology

Despite the obvious similarities between previous studies and those of Ruggles, Dunn, and Kaysen with respect to the deficiencies of a decentralized statistical system, several developments during the decades of the fifties and sixties did add new dimensions to these concerns. The first of these was the evolution of computerized data processing systems.

The federal statistical system, and in particular the Census Bureau, have long been in the forefront of the development and use of sophisticated data processing equipment. The first punch cards for automatic tabulation were developed by Herman Hollerith for use in the 1890 census.³⁷ The first practical large scale electronic data processing (EDP) system evolved from the Univac I which was developed in the early fifties for use by Census.³⁸ The Univac program also resulted in the development of the first practical form of compact machine readable data storage -- magnetic tape.

By the mid-sixties EDP systems were using the third generation of computers with vastly improved processing speeds, memory sizes, and storage capacities. Computers found increasing use in the federal statistical system resulting in tremendous increases in efficiency and reductions in cost in terms of data sorting and tabulation. Processing times were reduced drastically -- operations formerly requiring

months were reduced to weeks or days. The processing of the 1960 population census was reduced from an operation that could have taken several years to one which only required a few months.

More significant however, was the fact that the increased speed of computers permitted new types of analysis that in the past would have been prohibitively expensive and time consuming. Computers could edit and check raw data faster and more carefully than was possible manually, thus greatly improving the quality of final data. New types of tabulations could be developed. The same basic raw data could be analyzed in a number of different ways. In the past bodies of data were collected for use in only one or two specific tabulations which became the final -- and only -- form of the data then made available to users. The computer made possible the use of the data for new analytic purposes -- in part because of the low marginal costs of further processing, and in part because compact storage on magnetic tape made it practical to retain data for reuse. A single magnetic tape, for example, could hold the information contained on over 100,000 punch cards in a fraction of the space. Thus archival storage of important raw data became practical. In the past, little information could be kept for long periods as room was needed for the inflow of new punch cards. The compactness of magnetic tape also made practical the interagency transfer of large amounts of data and the combining and simultaneous analysis of such data.

The computer also allowed for far more detailed analysis of data than was previously possible. For example, prior to the introduction of the computer, aggregated tabulations of tax returns were

used to estimate the impact which proposed changes in the tax law might have on total tax revenue. With the introduction of the computer a more detailed analysis was developed. A sample of 100,000 tax returns was obtained and a computer program developed to recompute each tax return individually according to the proposed revision of the law and thus show for the sample not only the cumulative impact of the taxes but the nature of any income redistributions.

Coincident with the development of increasingly sophisticated data processing capabilities, the increasing emphasis on social programs in the sixties generated an enormous new set of data requirements for the development, planning, and evaluation of these programs. Programs dealing with inter alia, poverty, education, health, and urban organization all required increasingly detailed time series records on sub system components of the total economy to efficiently address these issues. Policy makers, as Dunn noted, needed records "to understand adequately how people, households, regions, activities, enterprises and administrative units are functionally related and how they change over time."³⁹

The central problem, which was fundamentally different than it might have been twenty years ago, was one of records and not data. That is, in general all the data necessary for the development of these records was already extant in some part of the federal statistical system. What was required, as Ruggles, Dunn, and Kaysen noted, was the integration of this data in a manner that allowed the generation of the necessary records and analyses. Before the advent of the computer, any such integration and analysis efforts would have been

prohibitively expensive and time consuming; the more realistic course at that time was the creation of new data gathering operations directed at the compilation of the specific records required. The development of the computer provided the technological capability to rapidly integrate and process existing data, but there still existed formidable barriers within the statistical system against the effective utilization of that capacity. As noted first by Ruggles and later by Dunn and Kaysen, the principal problems were:

The Archival Problem. This is the problem that initially interested the Ruggles committee. The problem arises from the fact that statistical agencies are often oriented to the production of one or a few particular publications, and they themselves have little use for the raw data after the completion of their publications. As a result data that might have been used profitably by other agencies or users is often destroyed.

In the past old data was destroyed as a matter of necessity to make room for incoming data. But with compact magnetic tape storage, physical considerations of space are no longer a concern. And with increasing interest and need in time dependent analyses, the lack of a coherent and consistent policy of data preservation is a serious problem. There is no adequate mechanism to insure that agencies are following optimal policies with respect to the preservation of important information.

Problems of File Maintenance. Even when raw data is saved by various agencies, it is sometimes so poorly documented as to

preclude its usefulness for any but its original application. Again, operating agencies are concerned with achieving the results necessary for specific tabulations or given computations. In the course of deriving these results errors are often discovered regarding transcription or classification of the raw data. Most frequently these errors are patched on an ad-hoc basis, but doing so still leaves errors on the original tapes which are seldom corrected.

In addition, agencies often do not provide sufficient information on the layouts, classifications, and definitions contained in a tape. As a result, even for the agencies' own purposes it becomes difficult to go back after a few years to make use of the information unless it happens to be in the same classification and format employed for current processing.

The Reference Problem. One of the serious problems blocking the fullest utilization of the existing data in the federal statistical system is the absence of any clearly defined reference function for identifying the data and its location within the system. This stems from the decentralized nature of the statistical system and its production orientation. To the extent that the agencies attempt to provide occasional referencing assistance, the task usually falls to a single individual whose existence frequently is not very well known outside the agency itself.

No agency is in a position to perform a reference service with respect to the total federal statistical file, and as demand

for increasingly complex and diverse types of data analysis grows, this becomes a more severe problem.

The Disclosure Problem. The legal and administrative regulations on the disclosure of information provided by individual respondents are becoming increasingly restrictive to users. Much of the data gathered by federal agencies is done so under the assumption that it will not be released either to outsiders or other government agencies. Any violation of that understanding would serve to severely jeopardize the continued cooperation of the agencies' sources. However, researchers rarely require specific information about respondents per se. The problem arises from attempts to associate individual sets of data to determine interesting relationships. The need for identifying individual respondents in any combining of data files arises not from an interest in individual responses but for the sake of appropriately matching data at levels of aggregation necessary to create the statistical analyses required. In other words, to cite a very simple example, suppose an economist is interested in what multiple of their yearly income a family will spend on a home, at various income levels. Further suppose that agency A has a tape that contains, for all of the people in a given city, names, addresses and yearly incomes. Agency B has a tape that has only addresses, names, and values of the homes at those addresses for the same city. The economist has no interest in either the values of individual homes or the incomes of particular

individuals. Yet the only way he can derive the statistical sample he wants is to match, at the family level, incomes with house values. If agencies are restricted by disclosure rules from providing this type of information to researchers, or each other, valuable correlations cannot be made.

Problems of File Compatibility. Related to the interagency utilization of data is the fact that classification and formatting schemes vary vastly between different agencies. This is again the product of the production orientation of the individual agencies which results in their tailoring their classification procedures in a manner that facilitates their production function. Unfortunately this tends to complicate any interagency use of the data, again, often preventing the fullest utilization of existing data.

Prior to the development of the computer, interagency use of data was simply not a consideration. Data was too bulky and the processing costs too high for it to be used by more than one agency. But the development of EDP techniques put a premium on the development of consistent interagency classification procedures.

These issues are, in brief, what Ruggles, Dunn, and Kaysen saw as the major impediments to the realization of the enormous potential computerized data processing had brought to the federal statistical system. But agreement on the nature of the problems did not imply agreement on the nature of the solutions.

The National Data Center

First to address the problems outlined above had been the Social Science Research Council committee, chaired by Dr. Richard Ruggles of Yale University, on the Preservation and Use of Economic Data. The other members of the committee, chosen by Ruggles, were Edwin Kuh of M.I.T., Stanley Lebergott of Wesleyan University, Guy Orcutt of the University of Wisconsin, and Joseph Peckman of the Brookings Institute. The committee's report, submitted in April of 1965 and referred to as the Ruggles report, in part because of his chairmanship of the committee and in part because he was the principal contributor, contained three recommendations, the first and most consequential being the call for the establishment of a National Data Center.⁴⁰ Four requirements for such a center were outlined.⁴¹

The Need for Interagency Authority. The National Data Center would have the authority to obtain any computer tapes produced by other federal agencies. The center would obtain duplicate copies from the agencies of clean, edited tapes, thus both the center and the agency would have the basic data available. "In this connection the Federal Data Center should keep track of statistical projects underway in the federal government and make sure in advance that the budget for each project includes the proper provision for making clean, edited tapes and providing the necessary accompanying information on classification and programming."

The Need for Computer Facilities. Ruggles saw the data center as not only furnishing basic information but also, in order to overcome interagency disclosure problems, performing production runs of specified analyses for researchers or other federal agencies (this service would be provided on a reimbursable basis). The center would necessarily therefore require substantial computer facilities.

The Need for Service Facilities. The data center would act as more than a repository and production facility. Its role was described by Ruggles as similar to that of the Library of Congress, providing systematic and comprehensive coverage of the material available in its area of competence. It would act to anticipate user needs so that the most useful information would be preserved and available in usable form and that unwanted information would not clog the system.

The Need for Administrative Arrangements. In order for the National Data Center to carry out its functions some new administrative arrangements would need to be made, as no existing agency had the authority to undertake all the tasks described. No specific suggestions were made regarding the nature of these arrangements except to note the limitations of existing agencies as potential homes for the data center.

The entire proposal was in fact principally a conceptual outline. Less than two pages of the fifteen-page report were devoted to an overview of the center. No mention was made of particular types of institu-

tional arrangements vis à vis other statistical agencies, the scope of the data to be contained in the center, or specific operational procedures to be followed by the center. What did emerge was Ruggles' philosophy that the key to the realization of the full potential of the statistical system lay in the centralization and integration to the maximum possible extent at the microlevel of all federal statistical files. That is, if all of the information on micro units (i.e. individuals, or business establishments, depending on the type of file) now contained in diverse files throughout the statistical system could be brought together into a single integrated file, this would vastly increase the range of analyses that could be performed on this information, thereby vastly increasing the utility of the federal statistical system.

The Ruggles committee, therefore, urged the Bureau of the Budget, as it had through the Division of Statistical Standards overall responsibility for coordinating the activities of the federal statistical system, to undertake the steps necessary to establish a National Data Center. The committee's views were sympathetically reviewed by Charles Schultze, director of BoB, himself an economist and a professional acquaintance of Ruggles, and formerly of the Brookings Institute. At the suggestion then of Schultze's office, Raymond Bowman, director of the Office of Statistical Standards and a career civil servant, took the next action on the data center proposal by commissioning an acquaintance of his, Edgar Dunn, to undertake a further study of the issues.

Another Perspective on the Data Center

Edgar Dunn, at the time he was appointed to review the data bank proposal, was a consultant with Resources for the Future, a Washington consulting firm. Prior to his position with RFF he had long been associated with the federal government and the federal statistical system. His greatest exposure to the problems of that system came as Deputy Assistant Secretary of Commerce for Economic Affairs; that office having had under its jurisdiction the coordination and review of programs for the Department of the Census and the Bureau of Economic Analysis. In his position Dr. Dunn was responsible for the initiation and implementation of several reforms in the statistical operations of those departments.

In his review of the Ruggles report, Dunn essentially agreed with the problems of the statistical system as outlined. Because of its greater length (46 pages vs. 15 for the Ruggles report) the review by Dunn was more detailed in its analysis but tended generally to support the work of Ruggles. Both Dunn and Ruggles favored some type of centralization of the statistical system to address the issues but differed on its exact nature. The emphasis in the Ruggles report was clearly on the construction of a central repository for federal statistical data, but in Dunn's view this failed to take into account some crucial considerations. While a centralized data repository could serve to eliminate interagency disclosure problems, develop consistent archival policies, provide easier access to data, etc. -- all important concerns -- the fundamental problem, which was

not automatically solved by simply bringing the data together, was that of file compatibility; that is, the standardization of formatting and classification schemes which would allow the successful integration of files from different sources. The key to this lies not necessarily in centralization of files but in

- 1) the development of administrative mechanisms to enforce the design and maintenance of all statistical files in accordance with a uniform set of standards, and in
- 2) the development of a technical system capability which would allow the reprocessing of these files in the variety of ways necessary to satisfy user requirements.

In Dunn's words

The effective provision of these services may require the assembly of some of the records into an integrated file, but this is defined by technical systems requirements and is not the central issue it is made to be by many representation.⁴²

In other words, once system-wide file compatibility was achieved, there would be, in principle, no need for physical centralization of data. A central processing facility, when called upon to develop a particular analysis or tabulation could simply obtain from the relevant agencies the data tapes required for the analysis and return the tapes on completion. In practice, some sets of data tapes would be required so frequently that it would be expedient for a central facility to maintain its own copies. Moreover, some tapes that might be of historical value to a particular class of users -- but not to the agency that generated them -- would also be kept in the archival

files of the central facility.

Dunn then went on to argue for the establishment of a National Data Service Center, the primary function of which would be to provide service to users of federal statistical data both inside and outside of government. The center would at the outset be designed to incorporate some basic functions⁴³

- 1) To direct the file storage and management for significant archival records in machine readable form for all participating agencies -- i.e., Dunn did envision a central repository for some critical statistical data.
- 2) To provide a central referral and reference source for the users of federal statistics.
- 3) To provide explicit facilitating services for the users of federal data -- e.g. file rearrangements and cross tabulations, tape duplications and translations, various forms of file modifications and special tabulations, etc. These services would be provided for those records under the direct management of the service center, and for records under the control of agencies whose operations or facilities did not allow them to adequately perform these services.
- 4) To develop computer hardware and software systems essential to the file management and servicing functions.
- 5) To provide staff support necessary to develop the standards necessary for system operation.
- 6) A research and analytic capability would be essential to the

success of these functions and would need to be developed at the center.

The Dunn report went well beyond the Ruggles document in proceeding to amplify on the scope and cost of some of these functions. The six functions listed above were classified into three areas and examined in some detail. First was the reference function. In reviewing the requirements for this function it was roughly estimated that it might take as much as five years to develop, at an annual average cost of 2 million dollars.⁴⁴ It was noted that this figure was a very crude estimate unsupported by any kind of extensive staff work that would be essential in a refined program plan.

The second major concern of the service center would be the collection of a core of statistical data tapes of considerable archival value under the direct control of the center. The question of which records constituted significant archives rested upon the interpretations of requirements and standards not yet developed. However, in order to get some feel for the scope of the problem Dunn asked Rudolph C. Mendelsohn of the Bureau of Labor Statistics to prepare a list of records considered to be a vital general purpose series.⁴⁵ On the basis of Mendelsohn's list, it was estimated that a complete archive might represent some 20,000 reels of magnetic tape and would require 3 to 3.5 million dollars to develop over three to five years. It was further estimated that nearly half of this file, or about 9,000 tapes, could, because of their relatively "clean" state, be brought directly into the archive for about \$260,000 within a year.

The 9,000 tapes would include inter alia the 750 reels of the Census Housing Data, the Census Current Population Data on 375 reels, the Bureau of Labor Statistics Consumer Expenditure Survey on 43 reels, the National Income Accounts on 2 reels, IRS tax data on 5,300 reels, and Social Security data on 1,900 reels.⁴⁶

The third area of concern was the service or facilitating function, that is, the various types of special tabulations and analyses that the center would perform for users. Again, the report took pains to note the speculative nature of the cost estimates which it placed in the neighborhood of 1 to 2 million dollars annually for the first several years of the center's operation, gradually rising to about 10 million dollars annually in the later years of the center's operation.⁴⁷

Dunn ended his report with two recommendations for immediate actions which might be taken under existing authority by the Budget Bureau:⁴⁸

- 1) Standards should be developed that would shape the content of archival records and determine the essential forms of file maintenance and documentation as part of an on-going program. A beginning should also be made in formulating the kinds of standards that would produce statistical building blocks essential to file compatibility.
- 2) The 9,000 tape file record identified in the Mendelssohn report constitutes a nucleus archive that should be generated quickly at a very modest cost. Funds should be made available

to the agencies to begin the creation of this basic archival record.

Dunn's second recommendation proved problematic in a number of ways. First of all, it tended to obscure the central distinction between Ruggles' views and his own with respect to the data center. Dunn had argued throughout his report that the key to realizing the full potential of the statistical system lay in assurance of file compatibility, not, as Ruggles had suggested, in massive file centralization. Yet Dunn's call for the immediate creation of a 9,000 tape data bank to act as a core for an even larger data center seemed very much like the massive file centralization Ruggles was advocating. While no one knew exactly how many data tapes existed throughout the federal statistical system (estimates had placed it in the range of 30,000),⁴⁹ 9,000 certainly represented a substantial fraction of the total stock, and an eventual file of 20,000 or more tapes sounded like much more than a limited archival file.

Dunn later explained that neither the 9,000 nor 20,000 figures was intended to be taken as anything more than very crude estimates of what the size of a potential archival file might be. The 9,000 tapes identified for initial centralization were chosen less because they were presumed to have considerable archival value than because they were relatively clean, edited tapes which could be brought together at comparatively low cost. The desire on Dunn's part was to quickly collect these tapes before they were destroyed by the originating agencies, then at some later date review them for their

archival value.⁵⁰ This point as Dunn admitted was never made clear in his report, and as a consequence, Dunn came to be identified along with Ruggles as a proponent of a data center.

What also later proved troublesome was Dunn's observation that the relatively low cost of this initial file meant it could probably be developed under existing authority with BoB contingency funds, which is to say without prior Congressional approval. Not surprisingly this proved to be a sore point in the subsequent Congressional hearings.

As it happened, Congress need not have been concerned that the Budget Bureau would be busily building up a data center behind its back. When the Dunn report was submitted to the Office of Statistical Standards in December of 1965, there was no move to implement either of Dunn's recommendations for immediate action. For the most part this was a consequence of the naturally cautious nature of Ray Bowman, then head of the Office of Statistical Standards. Bowman had displayed no particular interest in the data center. He had commissioned the rather small Dunn study (a total cost of less than \$12,000)⁵¹ at the suggestion of Charles Schultze's office, but took no part in it and had not interacted in any significant way with Dunn or Ruggles. Bowman was also clearly not interested in taking any initiative on what might be a potentially controversial issue, and he was aware that the one statistical agency that had shown any interest in the data center -- the Census Bureau -- was expressing serious reservations as to its utility. In particular, Morris Hansen, the chief statistician at Census was voicing his own opposition to a large scale data center.⁵² Hansen

felt massive data centralization and integration was both unnecessary and inefficient. His argument was that the process of matching two or more statistical files at the micro level, even if all employed the same classification and formatting schemes, was far from simple.⁵³ It was expensive, time consuming, and involved several important technical tradeoffs which were not adequately appreciated by Ruggles and other data bank proponents who had not worked as extensively with the manipulation of massive data files as had Hansen and other Census Bureau professionals.

Hansen's opposition might also have been influenced by organizational considerations. Other federal statistical agencies were unwilling to show much interest in the data center concept until a concrete proposal emerged defining the organizational relationships between it and them. Census, however, was the preeminent statistical agency and any massive centralization of statistical files not under its direct control would certainly detract from the Bureau's standing. It was in these terms that Ruggles saw Hansen's opposition, and it was for this and other reasons that Ruggles persuaded Charles Schultze to commission yet a third study of the statistical system.

The Kaysen Report and Its Implications

The Task Force on the Storage of and Access to U.S. Government Statistics was appointed by Charles Schultze and chaired by Carl Kaysen, then a professor of political economy and associate dean of the Graduate School of Public Administration at Harvard. The appointment of the task force and Dr. Kaysen's chairmanship were, at least partially if

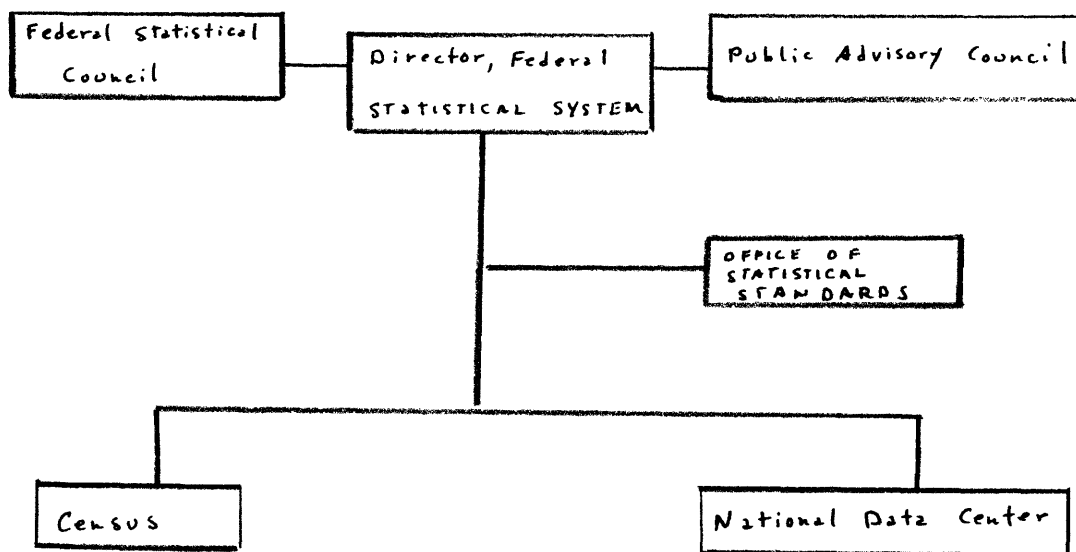
not primarily, due to the urgings of Dr. Ruggles. Ruggles himself served on the task force and was the principal author of its report.⁵⁴

For Ruggles the task force represented an opportunity to both clarify and gain support for his views against those of Dunn and Hansen. Not surprisingly then, the Kaysen report strongly recommended the creation of a National Data Center -- again using Ruggles' terminology -- having the following three principal functions.⁵⁵

- 1) The assembling in a single facility of all large scale systematic bodies of demographic, economic, and social data generated by the present data collection or administrative process of the federal government.
- 2) The integration of the data to the maximum feasible extent, and in such a way as to preserve as much as possible of the original information content of the whole body of records.
- 3) The provision of ready access to the information, within the laws governing disclosure, to all users in the government and, where appropriate, to qualified users outside the government on suitably compensatory terms.

From an organizational standpoint, the task force suggested that an entirely new position be created within the executive branch, that of Director of the Federal Statistical System, with the placement of the Census Bureau and the National Data Center as coordinate units under his direction. Other federal statistical agencies would be represented in the director's office through the Federal Statistical Council which would act as an advisory body to the director as would

a Public Advisory Council representing interests outside the federal government. As the new director would exercise coordinating powers over the federal statistical system, the Office of Statistical Standards would accordingly be transferred from the Bureau of the Budget to become a staff office of the new director. The organizational relationships of these elements are shown below.⁵⁶



The report was shorter and less detailed than that of Dunn in setting out what specific kinds of data might be included in the center and in exact functions it might serve. However, the report's emphasis on the centralization and integration of all large scale statistical files places it clearly at odds with the views of Dunn and Hansen. Moreover, its proposals for restructuring the federal statistical system -- with the data center on an equal footing with the Census Bureau -- were surely likely to provoke some opposition from Census.

As it happened, the Kaysen report received very little response. By the time it was published in the fall of 1966, the data center was no longer a viable issue. Nonetheless it is interesting to speculate on why the report took the position it did in seeking neither to accommodate the technical objections of Dunn or Hansen or to mollify Census' anxiety over the potential loss of status in the statistical system. What seems to emerge as the underlying motivation was the desire on the part of Ruggles and other members of the academic community to create within the government a centralized repository of statistical data that would be responsive primarily to the requirements of academic users. To maximize the center's responsiveness it would need to have both substantial data files of its own and the status of an independent agency. The lack of a substantial data collection, as would have been the case with Dunn's data service center, and the consequent dependency of the center on other agencies for data tapes would have resulted in these agencies inevitably exercising a greater degree of influence on the center than might have otherwise been possible. The center might thus become less responsive to the requirements of academic users and more concerned with agency issues. Similarly, the incorporation of the data center into any existing agency with operational responsibilities would have undermined the influence of the users community with the center. As for the technical arguments presented by Dunn and expressed by Hansen regarding centralization, they were quite simply ignored by the Kaysen report. The death of the data center issue prior to the release of the Kaysen report obviated the necessity for a formal dialogue to resolve this question.

Prior to the Congressional Hearings

Briefly summarizing then, as the data bank issue came before Congressional scrutiny in mid-1966, it was still a concept very much in its embryonic stages. Details as to the content and scope of its activities were far from being developed fully, and the whole activity involved only a handful of people.

Representing the academic community and arguing for a large centralized repository of federal statistical data was Richard Ruggles and a few economists. Contending that only a small archival repository was required and that emphasis ought to be placed on the development of a service capability within the statistical system was Edgar Dunn, an independent consultant and formerly a member of the Commerce Department. The distinction between these views tended to be obscured for a variety of reasons, and both were seen by Congress and the press as advocates of a national data center.

Within the executive branch of the government the data center concept in general had the sympathy of Bureau of the Budget director Charles Schultze, but Schultze was far too occupied with other matters to lend much active support or pay much attention to the course of the issue. The director of the Office of Statistical Standards, Raymond Bowman, who would have direct responsibility for the development and implementation of a concept of this sort, was disinterested at best. He had on the suggestion of the BoB director's office commissioned the Dunn review of the Ruggles report, but apparently paid little attention to the completed document.

The only statistical agency to have shown any concern with the data center concept at this stage was the Census Bureau, where chief statistician Morris Hansen had expressed serious reservations as to the utility of a centralized data file.

CHAPTER IV: THE GALLAGHER HEARINGS

The Privacy Issue and the Gallagher Committee

The issues of efficiency and responsiveness that had been the focus of executive branch discussions of the data center proposal were replaced in Congressional discussions with concerns over the invasion of personal privacy. The Congressional interest in personal privacy had been a long-standing one. In the Senate, the Subcommittee on Constitutional Rights, chaired by Senator Sam Ervin, had for at least eight years been involved in protecting the privacy rights of government employees.⁵⁷ In large measure it was the disclosures of this committee and the ensuing publicity, along with books and articles in the popular press, that had made privacy a highly visible issue in the mid-sixties.⁵⁸ Equally conspicuous in their concern over government intrusions on personal privacy were the Special Subcommittee of the House Committee on Government Operations, chaired by Representative Cornelius E. Gallagher of New Jersey,⁵⁹ and the Senate Subcommittee on Administrative Practice and Procedure, led by Senator Edward V. Long of Missouri.⁶⁰

The data center question was first taken up by Congress in June of 1966 when Edgar Dunn testified before the Long Committee. While somewhat related, the issue was clearly peripheral to the main thrust of the Subcommittee's investigations which focused on invasions of privacy through electronic bugging and wiretapping techniques. Dunn alone testified regarding the data center and his testimony constituted

only a few hours out of the set of hearings which went on for five days. Nonetheless, it was the first Congressional exposure of an issue that had already begun to attract attention in the press.⁶¹

The next Congressional action was a series of hearings dealing exclusively with the data center issue in July of 1966 before Congressman Gallagher's Special Subcommittee of the Committee on Government Operations.⁶² Gallagher had recently concluded a series of hearings on federal agency violations of the privacy rights of potential employees and had now turned his attention to other issues which appeared to constitute potential invasions of personal privacy by the government.⁶³ The issue he focused on was what he described as "the increasing demand for a centralized facility, within the structure of the national government, into which would be poured information collected from various agencies from which computers could draw selected facts."⁶⁴ Gallagher expressed concern that such a data bank would be established "in the interest of economy and efficiency"⁶⁵ without adequate attention being given to the impact it might have on personal privacy. The examination of that potential impact was set as the goal of the committee.

The Subcommittee itself consisted of three members, Gallagher, Congressman Benjamin J. Rosenthal of New York, and Congressman Frank Horton of New York. It can be safely said that none of the three were favorably predisposed toward the data center issue. Some comments from Congressman Gallagher's opening remarks were noted above.⁶⁶ The following two excerpts from their remarks reflect the views of Rosenthal and Horton respectively:

My own personal reaction to the proposal for a National Data Center was, I suppose, similar to that of most citizens -- intense apprehension at the prospect of still more invasions of personal privacy. In so many areas technological progress is being secured at the expense of personal liberty. The projected National Data Center seems an almost too fitting symbol for that development.⁶⁷

...there is danger that computers, because they are machines, will treat us as machines. They can supply the facts and, in effect, direct us from birth to death. They can "pigeonhole" us as their tapes decree, selecting, within a narrow range, the schooling we get, the jobs we work at, the money we earn and even the girl we marry...It is not enough to say "It can't happen here"; our grandfathers said that about television.⁶⁸

The Witnesses and the Issues

Eight witnesses appeared during the three days of hearings. First to testify on July 26 was Vance Packard, identified as an author, sociologist, and lecturer. Mr. Packard's books include "The Hidden Persuaders," "The Status Seekers," "The Wastemakers," and the "Naked Society," all dealing in some fashion with individual rights and motivation in a technological society. Testifying after Packard was Charles Reich, professor of law at Yale University and an expert on the legal implications of the collection of dossiers.

On July 27, the witnesses for the executive branch, Bowman, Ruggles, and Dunn, appeared, with Bowman's testimony taking all of the Committee's morning session while Ruggles and Dunn testified together during the afternoon session.

On July 28, Paul Baran and Burton Squires, two computer experts from the Rand Corporation and the University of Illinois, respectively, discussed the penetrability of computer systems before the Committee. Following their testimony, Robert R. Gallati, director of the New York

State Identification and Intelligence System, described the operation of that system to the Committee.

With the exception of Gallati, whose testimony consisted primarily of a factual description of a system's operation, and the three executive branch witnesses, the four remaining witnesses called before the Gallagher Committee were unanimously critical of the data center proposal. Some of the issues raised during the testimony of these four men are outlined below.

- One of the major concerns of Packard, Reich, and Congressman Gallagher was the extent to which the proposed data center constituted a first step toward a dossier society. As Packard noted,

The trail of records we leave behind us is steadily increasing. For example, within the government we have birth certificates, tax returns, the value of homes, locations of our banks, to some extent the value of our debts, our parents' and children's income, etc. Units of state government have school records, grades, IQ, driving records, criminal records, property holdings, marriage licenses, etc. If we happened to have applied for credit, or security clearances, the private and government agencies may have records regarding our personality, employment history, to some extent even appraisals of our social life and our sex life. Hospitals have medical records and if we have ever moved, then the moving companies are likely to have made complete inventories of our personal possessions.⁶⁹

It was also noted that the advent of the computer had made it far easier to collect and maintain all of these records. More to the point with respect to a national data center, the computer had made it easier to centralize all these records, that is, to collect information from a variety of sources and store it all in one place, thus providing a complete profile on any one individual. The data

center proposal was seen as an important step toward this type of centralization and the erosion of personal liberties.

In the first place centralization in itself was seen as an invasion of privacy. Records that, when widely scattered throughout the government, were not seen as particularly offensive intrusions, were perceived as real threats to privacy when collected together. In part this was due to the feeling that there is security in decentralization; developing a profile on an individual is much harder when information is stored in a number of different places, but can be had at the touch of a button when stored in a single computer file. Another important consideration was that the accuracy of the information tends to deteriorate the further it gets from the original source. Raw data which conveys certain information in one context may when taken out of that context, convey an entirely different message, and in the interests of economy, data centers are likely only to take a limited amount of information from original sources without taking much of the vital supporting data. Thus a man who might have received two or three speeding tickets in his lifetime or perhaps failed to pay some parking fines may well be shown to have a criminal record in a data center file without any explanation of the charges. The data center report might be accurate, as far as it goes, but nonetheless misleading.

- Similarly, there is a tendency to attach far too much authority to data produced by computer and to fail to question sufficiently the basis for the information. Reich, for example, focused on many

of the evaluation forms he was frequently required to fill out on students seeking jobs or graduate school positions. When and if his recommendation or evaluation was one of a few items used in making an evaluation of a candidate, it would be carefully examined, the candidate given a chance to examine and, if necessary, rebut it, or Reich himself could be contacted and asked to elaborate or clarify some points. When it becomes part of a massive central file, there is a tendency to accept it -- along with everything else -- without review and without question. Moreover, the whole evaluation process may be secret, candidates may have no right to review or rebut information. Most distressing is that in a central file, the information -- an unfavorable review or evaluation -- is always there. If someone is called bad by somebody, then he is still called bad -- in the central file -- 10 years later. In a decentralized system that particular evaluation may never be sought 10 years after it is made. But in a centralized file, it is there, it is available, and therefore it is likely to be used.

- Another consideration frequently raised during the hearings centered on the extent to which the data center might facilitate the transfer of nominally human decisions and actions to the computer. In particular, both Packard and Reich as well as Congressman Gallagher and other congressmen at the hearings had read and commented upon an article in the Saturday Review of July 23, 1966 by John W. Macy, chairman of the U.S. Civil Service Commission. Macy's article extolled the virtues of the computer system installed at the Civil

Service Commission. It glowingly detailed how computers schedule examinations with respect to time and location, how they grade test papers, how they print out notifications of test results, stuff the letters into envelopes and address and mail the envelopes -- all virtually without human intervention. He further discussed how computers may be used to perform personnel searches finding the best possible men for given jobs by using the computers "perfect" memory to pick out the records of government personnel with the specific characteristics required for the job to be filled.

Mr. Macy was clearly quite pleased with his system. The Committee made it quite clear it was not. What the Committee saw was neither cost savings nor efficiency increases that may have accrued from the installation of computers in the Civil Service, but rather the progressive dehumanization of the system.

- With its two computer experts the Committee focused on the technical aspects of the protection of personal privacy in data bank centers. The two principal issues addressed were:

- 1) To what extent could software or other internal safeguards be programmed or built into a data center that would prevent the unauthorized disclosure of information on individuals? Both Baran and Squires identified a number of techniques that could be used to make a system difficult to penetrate. These are, for example, various types of cryptographic techniques that could be used to encode the information, as well as programming techniques that would preclude the disclosure of information on

individuals. There are, moreover, various types of monitoring programs that can be employed to foil a series of requests designed to obtain sample sizes small enough to be able to isolate information on individuals. But in the final analysis, any reasonably clever programmer with access to the computer can, with sufficient time, penetrate even the most sophisticated system.

2) The second method of penetration which concerned the Committee was access through telephone lines on a distributed (time sharing or remote terminal) facility. Again all of the above comments apply. As Baran, who had worked extensively in trying to develop secure military systems, noted, various kinds of techniques can be employed to make penetration more difficult, but not impossible. The best, he concluded, that could be done would be to make it so expensive to penetrate the system that it simply wouldn't be worth the price for anyone to obtain the information.

Briefly, then, these points represented some of the concerns expressed by the Gallagher Committee during the hearings.

Responses from the Executive Branch

Raymond Bowman, director of the Office of Statistical Standards, accompanied by Paul Kurger the assistant chief of that office, was the first member of the executive branch to testify before the Gallagher Committee. Bowman's introductory remarks before the Committee focused

primarily on a review of the history of the data center proposal and the increasing demand for statistical data. Interestingly, Bowman never strongly endorsed the data center concept; he only pointed out to the Committee that it did represent a way of improving the storage of and access to federal statistical data. He also briefly assured the Committee that a statistical center would not be interested in compiling dossiers on individuals, and that privacy could be protected in such a center using the same sort of administrative procedures that currently so well served the Census Bureau.

The questions directed at Mr. Bowman were in three general areas,

- 1) the cost effectiveness of centralization,
- 2) the nature of the technical and administrative safeguards regarding the disclosure of information on individuals, and
- 3) the content for the proposed data center.

With respect to the cost/benefit question, Bowman displayed some unfamiliarity with the cost estimates prepared by the Dunn group and admitted to coming unprepared to outline the full range of agencies or groups who might be interested in using the center. He did describe in general terms how a data center would facilitate the interagency transfer and utilization of information.

With respect to the disclosure issue, Bowman argued for legislative controls on the disclosure of information from the data center and heavy legal penalties for violations. This sort of system, he maintained, had been highly successful in preventing unauthorized disclosures within the Census Bureau, and he would, therefore, expect

it to be equally successful for a statistical center. Bowman also made one noteworthy response to Congressman Gallagher's question as to why neither the Ruggles report nor the Dunn report contained any mention of the privacy problem or the disclosure of information on individuals. In his words, "The only way I can explain this is that all of us who were working in this area sort of assumed that it really didn't have to be discussed. In other words, the statistical system has been so imbued with the notion that you do not reveal information about the individual -- it didn't get mentioned. I am very sorry."⁷⁰ Bowman's view here, shared also by Ruggles and Dunn, serves to explain a conspicuous absence in those reports of any mention of the privacy issue.

In terms of technical penetration of the system, Bowman maintained that the actual makeup of the system had not yet been defined sufficiently to determine what sort of precautions would have to be taken to prevent this. What Bowman tried to impress upon the Committee, with little success, was that the data bank concept was as yet in no way ready for presentation before Congress as a proposal requiring approval and funding -- and when it finally was, these issues would be addressed in greater detail.

With regard to the content of the data center, again Bowman tried to indicate that these questions had not yet been fully resolved, and it had not been decided what would or would not be contained in the center. Bowman did indicate that the center would contain only statistical information and not intelligence records. However, he was not able to define to the satisfaction of the Committee the dis-

inction between these types of records nor did he attempt to provide any examples of what specific types of data files would be included and which excluded. He did not appear, when questioned, to be familiar with the 9,000-tape initial inventory proposed in the Dunn report.

Following Bowman, Ruggles and Dunn testified before the Committee. Ruggles' testimony dealt in large part with explaining to the Committee the necessity for retaining within a statistical file individual identification. In other words, why could names or other identifying numbers not be removed from the data once it was entered into a file. In the view of many of the congressmen, the failure to remove these names made a statistical file into an intelligence file. This is, of course, a very fundamental point which Ruggles took great pains to explain -- the need for some sort of individual identification was to enable the successful matching of files from different sources. For example, if one is to combine in any meaningful fashion, say IRS and Census data, one must be able to associate elements from each file for the same individual. While the Committee was finally willing to accept the necessity for individual identification in a statistical file, Congressman Gallagher took pains to point out the threats this entailed to personal privacy.

Dunn's testimony was more broad ranging, touching on most of the points covered by Bowman. There were some interesting contrasts between their approaches, however. Dunn in his opening statement took pains to distinguish between statistical information systems and intelligence systems in terms of their output. He then

turned to the protection of individual data, distinguishing between administrative and statutory protections and technical protections. He also considered some types of data which might be included in a statistical data center.

Elaborating in his testimony on some of these issues, Dunn pointed out that technical penetration of a statistical system to obtain individual data would always be possible. However, the effort involved in penetrating a properly designed system would most likely be quite prohibitive, and from a potential penetrator's point of view would probably cause him to look elsewhere to obtain the data. In terms of administrative safeguards, Dunn like Bowman referred frequently to the success of the Census Bureau system in preventing any disclosures of individual information.

Dunn also tried to clarify his position with regard to the 9,000-tape archive suggested in his report. In the view of the Committee this constituted only the beginning of a much more massive data file. Dunn tried to emphasize that the 9,000 tapes, suggested as an initial stock for the center, were tapes which were seen as having considerable archival value and ought not to be destroyed without careful review. The tapes did not represent a beginning for the data center, but were only those tapes which some members of his staff felt might be destroyed by the agencies which had them, with an attendant loss of valuable information. The data center would take charge of these tapes, review them for their archival value, and then store those which met the center's requirements, destroying the remainder.

Comments on the Hearings

The most obvious and important observations that can be made regarding the Gallagher hearings is how they illustrated the different priorities, perspectives, and assumptions which the members of the executive and legislative branches brought to the evaluation of the same issue. In the executive branch, discussions of the data center concept focused primarily on the extent to which it might improve the efficiency of federal statistical operations. In the legislative branch, the overriding concern was with personal privacy. Not one of the witnesses called before the Gallagher Committee, from outside the executive branch, was asked to offer a view as to potential improvements in the statistical system that might be realized by the creation of a data center. What they did comment on was the variety of ways in which such a center could potentially violate personal privacy rights.

The executive branch members testifying before Gallagher were simply unprepared for this preoccupation with personal privacy. Further, two major sets of assumptions, implicit in their thinking, but clearly not in that of Congress, made it difficult to carry on a meaningful dialogue on the issue. The first of these was that the disclosure problem in a statistical data center would not be a major issue. The members of the federal statistical community, including Ruggles and Dunn, were well acquainted with the effectiveness of the administrative mechanisms used by the Census Bureau to prevent unauthorized disclosures of individual information. It was thus assumed

that similar mechanisms would be employed, and prove equally effective, in preventing disclosures in a data center. But Congress had a broader view of the federal bureaucracy -- the record of the Census Bureau with respect to the disclosure issue may have been excellent, but the record of the IRS was abismal. The professionals in the executive branch may have assumed that the most rigid safeguards and the best administrative procedures would be imposed on the data center. But the members of the Gallagher Committee, with perhaps less than the highest regard for the caliber of the typical civil servant, feared that the loosest rules and sloppiest enforcement procedures would come to dominate.

When the executive branch failed to present during the hearings an extensive, well defined series of measures for the prevention of unauthorized disclosures from a potential data center, it was interpreted by Congress as an indication of a lack of serious concern for personal privacy on the part of the data center advocates. The advocates, for their part, felt that their proposed center was still at too conceptual a stage to need to address detailed considerations of this sort. The Census example proved that effective mechanisms against unauthorized disclosure could be developed, and when the plans for a data center did become more formalized, it would only be necessary to adapt these mechanisms to the new center.

The second major assumption underlying the attitude of some executive branch members toward the privacy issue was the traditional distinction made in the federal data gathering community between statistical, intelligence, and administrative files. The Orwellian scenarios,

developed by some members of Congress, of computerized men and intimately detailed birth-to-death dossiers stemming from the proposed center seemed absurd to members of the statistical community who felt there was a sharp distinction between their records and those of dossier compiling agencies such as the FBI. Long traditions had developed to insure the separateness of these types of files and prevent any significant interchanges of data. The suggestion that they might in some way be merged in a statistical data center was anathema to the members of the statistical community. But the distinction between these types of files was never that clear to Congressman Gallagher and the members of his Committee, and as they lacked an intuitive grasp of these distinctions, they sought specific examples of the types of data that would and would not be included in a statistical center. The failure of the data center advocates to provide this type of information led Congress and ultimately the public to the suspicion that a so-called statistical center was only the first step toward the total centralization of all government files.

CHAPTER V: CONCLUSIONS

Aftermath of the Gallagher Hearings

Reaction to the Gallagher hearings in the press was both substantial and uniformly critical of the data center proposal. Editorials and articles appeared in the New York Times, the Washington Post, the Washington Daily News, the Wall Street Journal, the New York Times Magazine, the Los Angeles Times, and the Providence Journal among others. Reports also appeared in U.S. News and World Report, The Nation, and The Nation's Business and in numerous specialized professional journals and trade publications.⁷¹

Only the Wall Street Journal presented an article advocating the proposal, and that in turn was followed by an editorial arguing against it.⁷² Press reaction ranged from attacking the proposed center as just another phase in the continual erosion of personal privacy rights in a technological society to bitter attacks on the concept as the first vital step toward a total police state. In the Los Angeles Times, Barry Goldwater asserted that the data center would give the White House police state power⁷³ while the Chicago Sentinel warned that "the data bank envisaged could become a Gestapo-concept, furnishing by push-button from coast-to-coast a complete list of all Zionists, integrationists, pacifists, Seventh Day Adventists, or Roman Catholics."⁷⁴

Within the Budget Bureau whatever support there may have been for the data center proposals evaporated under Congressional pressure

and public criticism. Ray Bowman, never particularly enthusiastic over the concept, had no intention of defending it any further after it came under fire. Charles Schultze, who had at least been sympathetic initially, saw little point in pursuing the issue in light of the Congressional and public responses.

Within the executive office only the Census Bureau had taken an interest in the proposals, and it had subsequently expressed strong reservations as to their value. This left only Ruggles, Dunn, and a handful of economists still supporting the concept, and with Bowman uninterested and Schultze having dismissed it as politically unfeasible, they had no organizational entry into the government through which to pursue the issue in any substantive way. The Kaysen report, submitted in October of 1966, was tabled, and the entire issue more or less officially died when in March of 1967, Congressman Gallagher announced that director Schultze had informed him that no further action would be taken on the data center proposal until the consequences and problems of the establishment of such a center could be fully studied.

During 1967, a second series of hearing on the data center by Senator Long helped maintain Congressional and public pressure on the Budget Bureau, insuring that no progress be made on the concept.⁷⁵

Conclusions

In reviewing the course of the data center proposal, one interesting observation is that the concept suffered the overwhelming rejection that it did partly because it was brought under public and

Congressional scrutiny long before it was ready for such appraisal. If its advocates in the executive branch had had perhaps another year to examine the issue, they might have been able to resolve some of the finer internal points of the proposal and present to the Congress a unified, well supported concept. Additional time might also have allowed for a better preparation of the strategy to be used in approaching Congress with a data bank proposal given public attitudes towards such a concept. The testimony before Gallagher by executive branch officials and the Dunn, Ruggles, and later Kaysen reports, all focused primarily on the data center as a means to improve government efficiency. Very little consideration was given toward presenting this proposal in some form that would allay public concern over the privacy issue. The distinction between statistical and intelligence systems was never forcefully enough drawn in any of the presentations. In the Dunn and Ruggles reports, this is understandable; they were intended for use only by the in-group which implicitly made such distinctions and had a high regard for the privacy issue as a matter of course. But in the Gallagher hearings the privacy issue was badly mishandled. Rather than assuring Congress and the public that any establishment of such a center would be contingent on the development of adequate privacy safeguards and that assessing the privacy implications of a proposed center would be matters of the highest priority, the executive branch representatives chose instead to try to defend a concept which had not yet been fully developed.

Of course, whether the executive branch, given more time, would indeed have focused more attention on the privacy issue is itself in

question. There are sound reasons for speculating both ways. The effectiveness of the Census system for preventing unauthorized disclosures and the traditional distinctions between statistical and intelligence data files that were implicit in the thinking of the data center proponents might have resulted in a continued omission of detailed considerations of the privacy question. The Kaysen report, for example, considered the privacy question in a brief two-page appendix written only after the Gallagher hearings. On the other hand, the Census Bureau was becoming increasingly aware of the privacy question surrounding statistical data as Congress began to focus its scrutiny on the content of the census questionnaires. Had Census and the data center proponents been able to resolve their differences, there is the possibility that the Bureau would have sensitized them to these issues.

But even had the executive branch taken a better approach with the Committee than it did, there is doubt as to whether any data center proposal could have survived the attack of a Congressional committee such as Gallagher's. If the press is any reflection of popular sentiment, a great many people felt as Gallagher did that centralization was in itself undesirable, that a centralized data bank, no matter how well protected, offered the government too much temptation to misuse it. Neither legal nor technical safeguards were failsafe, and the centralization of files eliminated one of the best safeguards of personal privacy -- the bureaucracy. No matter how carefully it was explained, the centralized data file would still be, in the minds of many, just one step away from a personal dossier bank, an indelible

birth-to-death record of our lives whose very existence was in some way an affront to human dignity.

Against these sentiments the proponents of the data center could only argue that it would improve the efficiency of government statistical operations. However desirable such a goal might have been, it certainly carried less emotional impact than the images of "big brother" in Washington, raised by opponents of the proposal. Disassociating the "big brother" image from the statistical data center would probably have proven too great a task for the center's advocates. Even if the public and Congress could be persuaded that the initial data stored in such a facility was of a non-sensitive nature, the very existence of the center would be a temptation to fill it with progressively more sensitive information -- all in the name of efficiency -- until it did, in fact, come to be a dossier file. The statistical community believed that the long standing traditions against mixing intelligence and statistical data would prevent this, but neither the public nor Congress was intimately acquainted with these traditions. Perhaps more important, the statistical community itself failed to consider how these traditions might change when adapted to the new structure of a centralized system.

In the final analysis, therefore, even more time, greater unity in the executive branch, and a better presentation to Congress and the public would not have won approval for the data center given the attitudes to personal privacy prevalent at the time.

As a final note, I began this study to investigate what seemed to me as one of the earliest technology assessments. As it happened

there were no real technical questions of consequence in dispute. There was a general agreement that a computerized data center could indeed improve the efficiency of the statistical system. There was also a general agreement that any such system could be technically penetrated. But the real issue was balancing efficiency and the threat to personal privacy. This was not a technical problem but a political one. A proven technology had both a positive and negative potential. Through the political process it was decided that the latter outweighed the former and the technology was never implemented. In light of the current discussions on technological imperatives and the inevitability of technological progress, this stands as an interesting counterpoint.

FOOTNOTES

1. John McCarthy, "Information," Scientific American 215 (1966).
2. The SSRC was called upon by the AEA because of their previous experience with the statistical system, See infra 26. For another account of the history of the SSRC report see "The Report of the Committee on the Preservation and Use of Economic Data to the Social Science Research Council," April 1965 (Hereafter referred to as the Ruggles report) contained in U.S. Congress, 89th Congress, 2nd Session. Hearings before a Subcommittee of the Committee on Government Operations, House of Representatives, Congressman Cornelius E. Gallagher, New Jersey, Chairman, The Computer and Invasion of Privacy, July 26-28, 1966 (Hereafter referred to as Gallagher) Appendix 1, pp. 195-196.
3. See infra p. 40 for the remaining members.
4. Ruggles, op. cit.
5. See infra, p. 28 regarding the trend toward decentralization in the statistical system.
6. Ruggles, op. cit., pp. 200-202.
7. ibid., pp. 195.
8. See infra, pp. 42-43.
9. Statistical Evaluation Report No. 6, Review of a Proposal for a National Data Center (Hereafter referred to as Dunn) contained in Gallagher, op. cit., Appendix 2, pp. 254-295.
10. Dunn, op. cit., p. 257.
11. This important point is discussed in greater detail infra, p. 43.

12. Dunn, op. cit., p. 276.
13. Gallagher, op. cit., p. 111.
14. ibid., p. 110, see the interchange between Congressman Horton and Dunn.
15. See infra, p. 57 regarding the appointment of the Kaysen task force.
16. The other members of the task force were: Charles Holt of the University of Wisconsin, Richard Holton of Berkeley, George Kozmetsky of the University of Texas, and Russell Morrison of Standard Statistics Co.
17. U.S. Congress, 89th Congress, 2nd Session, Hearings before the Subcommittee on Administrative Practice and Procedure of the Committee on the Judiciary, United States Senate, Edward V. Long of Missouri, Chairman, Topic: Invasions of Privacy, June 7-9, 14, 16, 1966. (Hereafter referred to as Long).
18. Gallagher, op. cit.
19. See infra, p. 56
20. U.S. Congress, 88th Congress, 2nd Session, Hearings before a Subcommittee of the House Committee on Government Operations, Cornelius E. Gallagher, Chairman, Topic: Use of Polygraphs as Lie Detectors by the Federal Government, 1965.
21. Gallagher, op. cit., p. 2.
22. ibid., p. 3.
23. See infra, p. 58.
24. See infra, p. 69.
25. See infra, p. 65.

26. See infra, p. 71.
27. See infra, p. 54.
28. Report of the Task Force on the Storage of and Access to Government Statistics (Carl Kaysen, Chairman) Executive Office of the President, Bureau of the Budget. (U.S.G.P.O.: Washington D.C., October 1966). Hereafter referred to as Kaysen.
29. Long, op. cit.
30. Alan Westin, Computers, Communications, and the Public Interest (The Johns Hopkins Press, Baltimore, Maryland, 1971) p. 156.
31. See the account of this in Records, Computers and the Rights of Citizens, Report of the Secretary's Advisory Committee on Automated Personnel Data Systems (M.I.T. Press, Cambridge, Mass., 1973) pp. 5-7.
32. See infra, p. 69.
33. For a more thorough review of the history of statistical reform movements, from which the information here is drawn, see Paul Fedman, "Commissions on Statistics: Statistics on Commissions," The Report of the President's Commission on Statistics (Washington, D.C., 1971) p. 477-497.
34. Kaysen, op. cit., p. 3.
35. Report of the President's Commission on Statistics, op. cit., pp. 41-42.
36. See infra, p. 71, on the death of the data center.
37. Records, Computers and the Rights of Citizens, op. cit., pp. 196-197.
38. Computing Surveys vol. 1, no. 1 (1961) p. 1.
39. Dunn, op. cit., p. 263.

40. See supra, p. 13, for the other two recommendations.
41. Ruggles, op. cit., pp. 202-204.
42. Dunn, op. cit., p. 266.
43. ibid., pp. 267-268.
44. ibid., p. 268.
45. ibid., pp. 276-284.
46. ibid., p. 276.
47. ibid., p. 270.
48. ibid., pp. 272-273.
49. Ruggles, op. cit., p. 201.
50. Gallagher, op. cit., pp. 110-111.
51. ibid., p. 71.
52. Hansen's views were best expressed in an article he later wrote titled, "The Role and Feasibility of a National Data Bank, Based on Matched Records and Alternatives," in The Report of the President's Commission on Statistics, op. cit., p. 1-63.
53. Suppose, for example, you wanted to merge two files, one containing names and yearly incomes, the second containing the same names and the values of property owned. The composite file thus ought to have a person's name and both his yearly income and the value of the property he owned. But now suppose there happen to be two John Smiths in the files. Depending on the programming, the computer can either discard both names and all the data associated with them, or it can arbitrarily match the file of John Smith on tape A, with the first John Smith on tape B. The match may or may not turn out right.

If both files contain names and addresses, this might help, but

people do change addresses. People also change names, sometimes legally, sometimes by misspelling them on forms, and sometimes by using different variations of their name, e.g. Jim Smith, Jimmy Smith, Jimmy S. Smith, etc. may all be the same person. Now extend all these problems to 2,000 files instead of two and there are major difficulties in record matching.

54. The other members of the task force are listed in note 16.

55. Kaysen, op. cit., p. 17.

56. ibid., p. 23.

57. See, e.g. Staff of Subcommittee on Constitutional Rights, Senate Committee on the Judiciary, 87th Congress, 2nd Session, Wiretapping and Eavesdropping Summary -- Report of Hearings 1958-1961 (Comm. print 1962).

58. E.g. Vance Packard, The Naked Society (1964); Myron Brenton, The Privacy Invaders (1964).

59. See note 20.

60. E.g. U.S. Congress, 89th Congress, 1st Session, Hearings of Invasions of Privacy by Government Agencies Before the Subcommittee on Administrative Practice and Procedure of the Senate Committee on the Judiciary (Edward V. Long Chairman).

61. Long, op. cit.

62. Gallagher, op. cit.

63. Personal reasons may have played no part in the Congressman's interest in the privacy issue, but for the record, it is worth noting that Mr. Gallagher was sentenced to 2 years in prison and fined \$10,000 for income tax evasion in 1973.

64. Gallagher, op. cit., p. 2.
65. ibid., p. 3.
66. Supra, p. 18.
67. Gallagher, op. cit., p. 4.
68. ibid., p. 5.
69. ibid., p. 8.
70. ibid., p. 72.
71. For a listing of the press articles commenting on the data center issue, see Annette Harrison, "The Problems of Privacy in the Computer Age: An Annotated Bibliography" (The Rand Corp., Santa Monica, Cal. 1967) pp. 111-113.
72. Wall Street Journal (August 5, 1966).
73. Los Angeles Times (June 24, 1966).
74. Chicago Sentinel (August 18, 1966).
75. Long, op. cit.