An Online Information System for Aggregate State and Local Area Economic Data

Charles G. Renfro*
Center for Applied Economic Research, 451 Commerce Building, College of Business and Economics, University of Kentucky, Lexington, KY 40506

This article describes the design and development of the Kentucky Economic Information System. This system, which is the first of its particular type, is a public computer-based data storage, retrieval, display, and analysis facility that has been developed in order to make aggregate economic data pertaining primarily to the state of Kentucky readily available to and easily usable by a wide variety of people. The system is currently resident on two separate computer networks and is accessible via a telecommunications terminal over an ordinary telephone line, as well as through local modes of access. The system consists of three integrated components: an extensive database; software to permit storage, retrieval, analysis, and display of the data; and a user manual which not only explains how to use the system, but also provides detailed descriptions of the data, in terms of concepts, collection or construction methodology, and reliability. This article discusses the various considerations that underlie the design of the system, describes the system's present characteristics, briefly evaluates the development of the system, and points out how this system provides a prototype for the development of other economic database systems.

The Kentucky Economic Information System, which is the subject of this article, is an economic database system; more generally, it is a type of numeric database system. The most recent, comprehensive review of numeric databases and systems is that by Luedke, Kovacs, and Fried [1]. They suggest certain defining attributes or characteristics of such systems. Specifically, they suggest that, in addition to providing one or more accessible databases, a numeric database system should possess two user attributes in particular: the abilities "to (1) directly locate and retrieve numeric data and (2) further manipulate the retrieved data by means of computational routines" [1, p. 120]. Essentially, they define a numeric database system as being "a complete system which can directly locate, retrieve, and further process and analyze numeric data" [1, pp. 120-121].

*Director, Center for Applied Economic Research, University of Kentucky, and Executive Director, Kentucky Council of Economic Advisors.

Received December 4, 1979; accepted March 21, 1980.

In the case of an economic database system, especially one containing aggregate, time series data, such capabilities are particularly important. These numeric data—which include such generally familiar measurements as Gross National Product, the unemployment rate, and the Consumer Price Index—are of relatively little interest in themselves. As a rule, it is upon transformation or in relation to other economic data, or under both circumstances, that particular economic measurements become meaningful. Furthermore, economists, who traditionally are the principal users of such data, will often wish to subject the data not only to simple algebraic manipulation, but also to various forms of statistical analysis; sometimes they will wish to create and use large-scale econometric models.

Economic database systems possessing the capability to support the construction of econometric models, as well as simple data retrieval and manipulation, have been slow to develop and proliferate. Various stand-alone statistical analysis packages have been available since at least the late 1960s and Statistics Canada, for example, has operated a countrywide, computer-based, economic data retrieval service, CANSIM, since 1966 [2,3]. However, it is only during the 1970s that systems have been developed that do not require the user to have considerable computer-programming experience to operate them effectively, yet which allow direct data location and retrieval for statistical analysis and econometric model construction and use.

In the United States, at a national level, such systems are offered to subscribers by the major econometric forecasting services: Chase Econometric Associates, Data Resources, and Wharton Econometric Forecasting Associates [1,4,5]. However, at the subnational level, the only system of this type is the Kentucky Economic Information System (KEIS) [6-8]. The KEIS is contemporaneous in its development with these national systems and incorporates, in its statistical analysis software, an earlier interactive system that was developed by the author in 1969-1970 to support the construction of the Brookings Quarterly Econometric Model of the United States. This system, then mounted on a PDP-10, was used generally by economists at the Brookings Institution during the period 1970-1971. The database systems now used by the three national econometric forecasting services began to be developed during this same period; In-

0002-8231/80/0031-0319$01.00

triligator [9] describes the relationships between the several national models.

The ability to support the construction and use of large-scale econometric models is an important aspect of the KEIS, but it is only one aspect. Possibly the most notable characteristic of the system is that it has been developed for general use. Unlike the national economic database systems just referred to, which are designed to serve the needs of economists and econometricians primarily, the KEIS has been developed in order to satisfy the requirements of a diverse group of users. As it exists today, the KEIS is a public facility that has been developed at the Center for Applied Economic Research, University of Kentucky, under the sponsorship of the Kentucky Council of Economic Advisors, a state government agency. The original purpose in the creation of the state database in 1973 was to support the construction of econometric models of Kentucky [10]. However, in 1976, in response to state government and other interest, the KEIS was made publicly accessible. A fundamental purpose of the system is to bring together the obtainable aggregate economic data for Kentucky in such a way as to make them readily available to and easily usable by government officials, businessmen, economists, and others [11].

The KEIS supports not only individual use but also the development of a number of projects, most of which are independently managed. For example, at the state level, it continues to support the development and use of econometric and other forecasting and simulation models [10-13]. At the substate level, it is beginning to be used for such purposes as the construction of an econometric model of the city of Louisville. The database includes both public and private segments—as will be described—and because of this, projects have been established at a number of universities in Kentucky, using the KEIS software and at least some of its public data, in order to sponsor quantitative economic research such as the Louisville econometric model and the study of particular Appalachian counties. In addition, data retrieved from the system and plots and other displays produced by it are featured in various regular and occasional publications of state government agencies and other organizations [13-16]. Furthermore, a subset of the KEIS and some of its data will soon become available to the large number of people who are to be served by the Agricultural Network Serving Education and Research (ANSER) and the Development Information System for Kentucky (DISK), related projects that are using minicomputer technology in conjunction with an IBM 370 mainframe [17]. While these projects are independent of the KEIS in most respects, including management and development, they are expected to provide some KEIS data and facilities to various local government officials and Agricultural Extension personnel throughout the state, among others. Finally, although the KEIS has not been developed particularly to serve the general public, it is available at this level through services provided by some university libraries in Kentucky; more generally, the Kentucky Department of Libraries and Archives has indicated its intention to make the KEIS available to 113 public library systems throughout Kentucky using their library network system, KENCLIP.

In order to serve this variety of users, the KEIS is now accessible on two separate mainframe computer networks—the Kentucky Educational Computing Network (KECNET) and the Kentucky state government network. KECNET primarily serves the state-supported institutions of higher education in Kentucky. The state government network primarily serves the state government, certain federal government agency units, and some national organizations, such as the Council of State Governments. However, others can gain access, for each of the networks can be used via a telecommunications terminal over an ordinary voice-grade telephone line from anywhere, as well as from dedicated remote terminals throughout Kentucky. The computers used are an IBM 370/165 Mod II, located at the University of Kentucky, an IBM 370/168 Multiprocessor system, and an IBM 370/3033, the last two of which are located in Frankfort, Kentucky. Between the two networks, the KEIS is available to users in any of three possible modes: as a batch system, a remote-batch system, or an interactive, time-sharing system.

The General Significance of the KEIS

The Kentucky Economic Information System can be considered from at least two perspectives: in a regional context, as an economic database system that is specifically designed to support regional economic analysis and to provide data pertaining to a particular geographical area, namely the state of Kentucky. Seen in this context, it presents a case study of what can be done for a state. However, the KEIS is also more generally relevant: it can be regarded as being an example of an economic database system that has been designed for use by a wide range of people, and which, only incidentally, happens to have a publicly accessible database containing regional data. Actually, the KEIS database is not limited to regional data; it also contains a substantial amount of national economic data. Seen in this more general context, as an economic database system capable of supporting economic analysis without necessarily any geographical restrictions, the KEIS represents a particular approach to the provision of economic data, and analytic and modeling capabilities.

There are two ideas which stand behind and are fundamental to the development of the KEIS. The first is that of bringing together data from many sources, most of which data cannot be obtained elsewhere in machine-readable form and some of which are otherwise unpublished in any form. The second idea is that of providing a system which allows the convenient use of economic data by a variety of users at whatever level of sophistication suits each user individually. In fact, in the design of the KEIS, "convenient" has been given an interpretation which transcends the computer-based capabilities of the system. Making it convenient to use involves not only the software design, but also the provision of detailed information to the user about the data contained in the publicly accessible database.

The Design of the KEIS

The Kentucky Economic Information System essentially consists of three interrelated components:

- Computer-resident, online database
- Computer software to store, retrieve, manipulate, and display data
- User manual, to provide information about the data in the public database, as well as information on how to access and use the KEIS

The various projects and publications that are supported by the system represent applications of it, rather than necessarily constituent elements. However, it should be noted that whereas this definition of the extent of the KEIS may seem natural, it actually represents a particular interpretation of the term "economic information system," for this term has been used in various ways. At one extreme, an economic information system, so called, can consist of the contents of a few manila folders in a file cabinet. At another remove, as in the case of the Minnesota Energy-Economic Information System [18], for example, such a facility can consist of a forecasting model, with only a tenuous connection to a database and associated software.

The KEIS will be described in terms of its software, its public database, and its user manual. The focus will be upon the design and development of the system as a public system. The perspective will be of the system as an economic database system, rather than necessarily a regional economic database system. However, in the discussion of economic data and their characteristics, the emphasis will be upon state and local data. This type of data is, in fact, of greater general interest than are national economic data. It is, after all, in the context of a relatively small, regional economy that most people live and work, and to the extent that people—particularly businessmen—are interested in data for large areas, often the data they require are not national data but data referring to some nonstandard group of states or local areas.

Software Design

There are several things that should be taken into account in the design of the software for a database system. Clearly, it is important to consider the characteristics of the computers it will be mounted on and the operating systems it will be run under. However, the users must also be considered. The prospective use of an economic database system by a diverse group of users obviously imposes particular design requirements. The foremost of these is access, including ease of access as well as simple availability. Computerized economic data are unfamiliar to most people and therefore must be made easily accessible in order to encourage use. In the case of the KEIS, this requirement involved mounting the system on the appropriate computers; for example, an attraction of the state government computers was the existence of more than 500 dedicated terminals in state government offices in Kentucky. However, equally important—and more pertinent to software design—a public economic database system must be easy to use whenever it is accessed. Precisely what makes a system easy to use will, of course, depend on the audience, but a criterion which might be adopted is that the system should not impose upon a user the need for any substantial prior computer-use experience. Finally, there is the consideration that, for the system to be immediately acceptable, the data which it makes available and the work done by the computer software must satisfy a prospective user's particular needs.

The prime audience for the KEIS consists of government officials, academics, and some business economists. Generally, those most likely to want to use the system will have had some experience with economic data; often they will have used the computer before and some will be very experienced computer users. However, in order to encourage maximum use, the software has been designed so as to be at least marginally usable by people who have had no prior computer experience and whose technical background could range from little or none to considerable. Thus, there are KEIS programs which allow the simple retrieval of data and the nearly automatic production of tables and plots; there are programs which provide statistical analysis and model building and use capabilities, as discussed earlier; and there are programs that require a level of technical knowledge in between these two bounds. However, all programs have been designed for ease of operation. Generally, the input is free format and the programs share a common command language.

There are two principles of design which have been adopted. The first is that the programs should support learning-by-doing. It is debatable whether a database system, and its supporting documentation, should be designed to teach statistics and econometrics, but it should sponsor self-instruction in the use of the computer, and the system's features in particular. The second principle is that the KEIS should help the user to learn how to use economic data effectively; obviously, one of the purposes of the system is to foster the study of the characteristics of economic activity. Of course, using economic data effectively is not necessarily the same thing as using sophisticated statistical and econometric techniques.

There are several categories of KEIS programs and among the most basic are those which simply allow data to be retrieved and printed in various report formats or else displayed graphically. These programs tend to give the user relatively little control over his output and are particularly easy to use. A good example of such a program is one called TABLE, the execution of which causes the automatic printing of a number of tables of selected data. This program requires no initiative on the part of the user; he simply executes it, copying the input commands from the system user manual. Alternatively, the user can execute EISTAB, which is also a table-printing program, but which requires him to define the data content of each of the tables that are printed, although not the table format. At the next level, as the case of GENTAB, the user must specify both the data content of the tables and the particular characteristics of the table formats; this type of program provides much greater flexibility in use, but, as a consequence, GENTAB

obviously requires more computer experience to use than either TABLE or EISTAB.

However, it is also possible to progress from the simple table-printing program, TABLE, to another simple table-printing program, BEATAB, which also provides some basis for analysis of the data. BEATAB almost automatically produces a table showing aggregate Personal Income and its components, and it displays not only the annual values of the items for each of two years but also the percentage of total income each component represents and the year-to-year percentage change. Alternatively, there is another simple program, RANK, that retrieves county data and exhibits the ranking of the counties by value of item; for example, this program can be used to rank the counties in terms of relative population, per capita personal income, the number of automobiles registered, or any of a number of other characteristics. The KEIS programs thus progress in their level of sophistication in terms of both the user input requirements and the output which results, at a given input requirement level.

There are also KEIS programs that allow a user to progress simultaneously in levels of technical sophistication and user input requirements. For example, DTP, which can be used for simple data retrieval and for the printing of particular data observations, is also a general-purpose data transformation program (hence DTP) allowing a variety of transformations on directly retrieved data. Similarly, a companion program, DTPPLOT, permits not only the plotting of series against time or against another series, but also the transformation of data prior to plotting. Each of these programs supports a considerable range of sophistication in use.

At a more advanced level, among the statistical analysis programs, there is a simple least-squares regression program, OLS, that is elementary in its statistical operations. There is also a technically more sophisticated program, GLS, which allows various types of regression to be performed. However, these programs are designed so that the program control commands for OLS and the operations performed by it are a subset of those for GLS, thus allowing the user to begin simply and then progress. For model building, these programs are complemented by several others that can be used for model solution, testing, and simulation, as well as for such bookkeeping operations as creating a glossary of model variables automatically.

Provision is also made for users to retrieve data from the KEIS database and to load them into various standard, commercially available statistical analysis packages. Consequently, the user of the KEIS is not restricted to programs that are formally part of the system software. He can even use data from the KEIS with programs he supplies. Of course, a design benefit of allowing the export of KEIS data is that it removes the need to satisfy the unique requirements of individual users or to include in the KEIS software infrequently used analysis features that are available in commercial stand-alone packages.

Another, very important, KEIS software feature, which was referred to earlier, is a set of programs that allow the user to create and maintain his own private database as a separate repository of data. This database can contain data from various sources: it can contain data that are retrieved from the public KEIS database as well as those supplied by the user. This type of database can be accessed by the KEIS programs in almost exactly the same way as the public database; the only technical difference of any concern to the user is that he is able to put data into his private database as well as to retrieve from it. The KEIS software thus provides not only the ability to retrieve and use KEIS data, but also the general capability to store, retrieve, display, and analyze any time series data. Considered from this perspective, the KEIS programs form a personal data management and analysis system for each user who wishes to employ them in this role.

A further characteristic of the software has been implicit in the discussion: that it consists of a number of separate programs. As an alternative, it could have taken the form of a single, all-purpose computer program. Arguably, this alternative provides a particularly convenient structure for the expert user in the case of systems that are used only interactively. However, such a program is often relatively time-consuming to learn to operate, particularly for the user who is only interested in performing a few simple tasks. In addition, although it could be written in modular fashion, such a program can be difficult to modify, especially if the person attempting to modify it is not its author.

In contrast, with the possible exception of the case of an experienced user in an interactive environment, there are certain virtues in designing an economic database system as a group of programs. One is that in order to perform a particular operation, a small single-purpose program often costs less to operate than a large program, since it normally uses less computer resources. A single-purpose program is also generally easier to learn to use than a multipurpose program and normally can be written much faster than a large and complicated one. Furthermore, extending the user options provided by a multiprogram system will often involve simply adding further programs, not revising one written earlier.

The disadvantages of a multiprogram structure depend on how the programs are written. If all the programs are written strictly as single-purpose programs, then to perform a set of related operations on particular data may require the sequential execution of a large number of programs. If the user-specified operating commands differ radically between programs, then the system as a whole can become difficult to learn to operate. If the programs write files that are not compatible between programs, then the operation of the programs as a system can become nearly impossible.

The software design for the KEIS takes into account these considerations. Certain programs are single purpose, but others perform a variety of related operations. As far as possible, the programs share a common control structure and use the same program control commands for identical operations; internally, many of the subprograms—functions and subroutines—are interchangeable between programs,

yielding the ability to create new programs quickly from prefabricated structures. Furthermore, except as required by differences in the operating systems, executing the KEIS programs on one network, rather than the other, involves few if any changes in the program control commands.

Database Organization and Structure

There are two aspects of the development of the KEIS database which deserve particular attention: the focus of this database and its intended degree of comprehensiveness. As indicated earlier, the public database has been developed primarily in order to provide users of the system with aggregate, time series economic data pertaining to the state of Kentucky. However, what makes this characteristic of the system especially noteworthy is not that the data made available tend to refer to a specific geographically defined area, nor that they are aggregate, time series data; to some degree, the KEIS shares both these properties with other economic database systems. It is that the public database is being developed systematically in order to provide a comprehensive collection of economic data with such characteristics.

There are certain restrictions on the data stored in the public system. These data cannot be confidential in nature nor can they refer to individual persons. For a time series to be included, the estimates must be made with reference to a particular conceptual definition of what the measurements represent, and the methodology used to make these estimates must be reasonable, when judged against professional standards. The effect of these provisions should not be underestimated, for they significantly restrict the availability of economic data for inclusion in the KEIS. However, except for data that are excluded for one of these reasons, the intention in the development of the public database is to provide ultimately a complete machine-readable source of aggregate, time series economic data relating to Kentucky.

Problems of Database Organization: It might appear that, in principal at least, collecting and organizing economic data pertaining to Kentucky would be a relatively simple task. In fact, the systematic provision of economic data pertaining to any geographic area involves vexing problems of definition. At the most general level, one of these is defining the term "economic data." It is possible, of course, to restrict the use of this term to numbers generated as the direct result of an attempt to measure specific abstract economic quantities. However, not only do many economic data series originate as a by-product of the administrative records compiled by various government agencies, but economists in their empirical work tend to accept as economic quantities a broad range of variables; for example, meteorological data and other physical measurements may be justifiably classified as economic data in some instances, as may various other measurements on phenomena that affect human life. In the selection of data to be included in an economic database, the issue is not that of defining exclusive categories of data; instead, it is a matter of identifying those data that will be of serious interest to users of that database.

The nature of the particular geographic area as an economy must also be considered. A state, for example, is not an isolated economic entity but shares institutional characteristics with all other states. There are national data that refer as much to individual states as to the country as a whole; as a case in point, consider federal tax rates. More generally, the behavior of one state economy may both affect and be affected by the behavior of other state economies; thus it may be appropriate to include data that ostensibly refer to these other economies. It may even be appropriate to include data on foreign countries, since states can be involved in foreign trade and are affected by world economic events.

However, a state is not an indivisible economic unit. States are composed of a number of different administrative and geographical subunits. The immediate question is, should the database contain all the available aggregate economic data for all possible state subdivisions? Such subdivisions include cities, counties, congressional districts, school districts, water districts, and the like. They also include such areas as Census tracts and enumeration districts, the latter defined so as to represent a reasonable workload for an enumerator.

It is evident that aggregate, time series economic data referring to a particular state do not necessarily form a well-defined, easily organized group. However, geographical representation only just begins to define the potential extent of an economic database, even given a strict definition of economic data. Aggregate economic data can also be classified in terms of their frequency of observation, types of economic activities covered, and their possible levels of aggregation. At minimum, a four-way classification of economic data is needed: Geographic Reference, Frequency of Observation, Specific Economic Activity, and Level of Aggregation.

Even these attributes together do not fully quantify the amount of data in a particular economic database at any given point in time. One of the reasons they do not is that inevitably some observations will be missing. However, it is also possible to classify economic data in terms of any prior transformations that might have been made; for example, quarterly and monthly observations are often released by original data sources in both seasonally adjusted and unadjusted form. In short, any simple classification scheme that is adopted for economic data may be subject to further refinement.

These issues are considered in greater detail elsewhere [11,19], but enough has been said to indicate the type of definitional and organizational problems that are involved. With these problems in mind, certain normative criteria were accepted in order to guide the organization of the KEIS public database: an economic database should be organized efficiently, it should be comprehensible to users, and it should be expansible without reorganization. In this context, efficiency should be understood to refer not only to the mechanics of data retrieval but also to the degree of

redundancy in the storage of economic data. Comprehensibility refers to the ability of users to find particular data series and understand their specific characteristics. As far as expansion is concerned, it should not involve database reorganization—the addition of one or more new data series should not affect the use of series previously included in the database. The structural characteristics of the KEIS database reflect the application of these criteria.

The Structural Characteristics of the KEIS Database: The KEIS database is organized into *data banks* and each of these can be visualized as consisting of a particular set of economic data series and associated descriptive information. A given data bank will be classified as either a public or a private bank. As indicated earlier, the public banks by definition are those that have been created and are maintained by the KEIS project staff for the use of anyone who accesses the system; together, the public banks form the *public database.* The private banks, of course, are those created by individual system users; a set of private banks created by a particular user is called a *private database.* However, although logically a part of the KEIS, private databases are physically distinct from the public database.

Irrespective of the type of bank, the data are stored in random access form and each data series is allocated two records, which together contain the observations on economic quantities and certain descriptive information. One of these records, the *data record,* includes a name—usually one to eight characters in length—which identifies the specific data series, as well as the initial and ending dates of series availability, a number identifying the periodicity of the data, and a vector of observations. The other record, the *documentation record,* contains the series name, a description (maximum of 72 alphanumeric characters in length), code word descriptors of the source of the data and of the units in which the observations are stored, and the initial and ending dates of series availability. Actually, each of these records will also contain other information about the data series, but this account is sufficient to convey an idea of the record content.

Currently, each of the public banks can be defined in terms of two of the characteristics of the data it contains: Geographic Reference and Frequency of Observation. There are now four public banks, and three of these contain data that refer to the state as a whole. The three are known collectively as the *State Banks.* The fourth, known as the *County Bank,* contains data on counties and various aggregates of counties.

The State Banks are distinguished from each other by the frequency of the data they contain, and they consist of the *Monthly Bank,* the *Quarterly Bank,* and the *Annual Bank.* Each of the data series in these banks begins in 1951, or the earliest subsequent date of first availability, and ends with the latest observation obtainable. In contrast, the County Bank contains only annual data and each series begins in 1950, or the earliest subsequent date of availability. In terms of the number of series, the County Bank is the largest; there are approximately 50 series per county and each of several county aggregates, for a total of more than 9000 series. The Annual Bank is the next largest with more than 4000 series. The Quarterly Bank has just over 750 series and the Monthly nearly 760. Together, these banks contain something in excess of a half million observations.

Associated with each bank is a computer-printed index that is designed to serve two purposes: (1) to provide the user of the KEIS with a means of determining the contents of the bank and, hence, of locating particular data series; and (2) to give a short description of each data series. Table 1 illustrates a portion of the Annual Bank index. Note that each line in the body of the index refers to a specific data series and that it includes a brief description of the series, as well as identifying the original source, the units in which the observations are stored, the period of availability, and other such information. As a by-product, the bank indices also form a checklist of the existing data on economic activity in Kentucky, organized by data bank. At the moment, these indices constitute only a partial list, but progressively the list will become more complete.

It is immediately evident that this logical organization of the database means that the user must know which bank a particular series is in, or else spend time searching for it across several banks. For some people, it is possible that this search could prove troublesome. However, while in principle the KEIS could be accessed by a person with no experience of economic data, in practice users are above this level. Almost all appear to be able to identify instantly the geographic reference of the data they seek. To the extent that a few people find it difficult to discriminate between banks, the problems they have seem to be in determining frequencies of observation; a user seeking a certain type of data does not always know the frequencies at which these data are available, nor if they can be obtained from a reliable source at any frequency.

In order to reduce the effect of this frequency determination problem, the State Banks are cross-referenced. Since data that are obtained originally as monthly or quarterly observations are routinely cumulated to an annual frequency and stored in the Annual Bank, this cross-referencing allows the Annual Bank index to serve as a global index to all the State Banks and, therefore, as a finding list for state-level economic data. In Table 1, the columns labeled "QRT" and "MON" indicate the series numbers in the Quarterly and Monthly Banks of those corresponding series that are available at each of these frequencies; the lack of an entry in either of these columns indicates the absence of a corresponding series at that particular frequency.

That the organization of the public database into data banks helps to achieve the objectives of efficiency, expansibility, and comprehensibility is almost self-evident. Narrowly defined, "efficiency" refers to finding and retrieving data series. All other things equal, the subdivision of the database into data banks obviously tends to increase efficiency in this sense. More generally, considering the question of redundancy, there is no organizational need for any duplication in the storage of individual data series because the KEIS classification scheme defines a particular location in the database for a specific series. In fact, some duplica-

TABLE 1. Kentucky Economic Information System: Index for the Kentucky Annual Economic Data Bank as of 16 April 1979.

SERIES NUMBER	NAME	DESCRIPTION	QRT	MON	UNITS	SOURCE	AVAILABLE
368	WAAAPP$	EARNINGS, AT ANNUAL RATE, APPAREL (SIC 23)	323	93	M$	DHR-EIS	1956-1978
369	WAAMBS$	EARNINGS, AT ANNUAL RATE, MENS' & BOYS' SUITS, COATS (SIC 231, 232)	324	94	M$	DHR-EIS	1956-1978
370	WAAPAP$	EARNINGS, AT ANNUAL RATE, PAPER & ALLIED PRODUCTS (SIC 26)	325	95	M$	DHR-EIS	1956-1978
371	WAAPRT$	EARNINGS, AT ANNUAL RATE, PRINTING & PUBLISHING (SIC 27)	326	96	M$	DHR-EIS	1958-1978
372	WAACHM$	EARNINGS, AT ANNUAL RATE, CHEMICALS & ALLIED PRODUCTS (SIC 28)	327	97	M$	DHR-EIS	1958-1978
373	WAAICM$	EARNINGS, AT ANNUAL RATE, INDUSTRIAL CHEMICALS (SIC 281)	328	98	M$	DHR-EIS	1958-1978
374	WAAPMS$	EARNINGS, AT ANNUAL RATE, PLASTIC MATERIALS & SYN (SIC 282)	329	99	M$	DHR-EIS	1958-1978
375	WAAPNT$	EARNINGS, AT ANNUAL RATE, PAINTS & ALLIED PRODUCTS (SIC 285)	330	100	M$	DHR-EIS	1958-1978
376	WAAPTRP$	EARNINGS, AT ANNUAL RATE, PETROL, COAL, RUBBER&PLAS PROD (SIC 29-30)	331	101	M$	DHR-EIS	1956-1978
377	WAAPET$	EARNINGS, AT ANNUAL RATE, PETRO REFINING & RELATED IND (SIC 29)	332	102	M$	DHR-EIS	1976-1978
378	WAARPP$	EARNINGS, AT ANNUAL RATE, RUBBER & MISC PLASTIC PDTS (SIC 30)	333	103	M$	DHR-EIS	1976-1978
379	WAALTH$	EARNINGS, AT ANNUAL RATE, LEATHER & LEATHER PRODUCTS (SIC 31)	334	104	M$	DHR-EIS	1956-1978
380	WAAMD$	EARNINGS, AT ANNUAL RATE, MANUFACTURING DURABLES	335	105	M$	DHR-EIS	1952-1978
381	WAALUM$	EARNINGS, AT ANNUAL RATE, LUMBER & WOOD PRODUCTS, EX FURN (SIC 24)	336	106	M$	DHR-EIS	1957-1978
382	WAAFUR$	EARNINGS, AT ANNUAL RATE, FURNITURE & FIXTURES (SIC 25)	337	107	M$	DHR-EIS	1957-1978
383	WAASCG$	EARNINGS, AT ANNUAL RATE, STONE, CLAY & GLASS (SIC 32)	338	108	M$	DHR-EIS	1952-1978
384	WAAPRI$	EARNINGS, AT ANNUAL RATE, PRIMARY METALS (SIC 33)	339	109	M$	DHR-EIS	1952-1978
385	WAABFS$	EARNINGS, AT ANNUAL RATE, BLAST FURNACE & BASIC STEEL PROD (SIC 331)	340	110	M$	DHR-EIS	1956-1978
386	WAAFAB$	EARNINGS, AT ANNUAL RATE, FABRICATED METAL PRODUCTS (SIC 34)	341	111	M$	DHR-EIS	1956-1978
387	WAAAMAC$	EARNINGS, AT ANNUAL RATE, ALL MACHINERY (SIC 35 & 36)	342	112	M$	DHR-EIS	1956-1978
388	WAAMAC$	EARNINGS, AT ANNUAL RATE, MACHINERY, EX ELECTRICAL (SIC 35)	343	113	M$	DHR-EIS	1958-1978
389	WAAELC$	EARNINGS, AT ANNUAL RATE, ELECTRICAL EQUIPMENT & SUPPLIES (SIC 36)	344	114	M$	DHR-EIS	1958-1978
390	WAATRA$	EARNINGS, AT ANNUAL RATE, TRANSPORTATION EQUIPMENT (SIC 37)	345	115	M$	DHR-EIS	1956-1978
391	WAAMDO$	EARNINGS, AT ANNUAL RATE, OTHER DURABLE GOODS (SIC 38.39)	346	116	M$	DHR-EIS	1952-1978
392	WAABCL$	EARNINGS, AT ANNUAL RATE, BITUMINOUS COAL & LIGNITE MINING (SIC 12)	347	117	M$	DHR-EIS	1973-1978
400	LFCIV	LABOR FORCE, TOTAL CIVILIAN	130	120	TPPL	DHR-CPS	1961-1978
401	LFPRATE	LABOR FORCE, PARTICIPATION RATE(LABOR FORCE/POP GTE 16)			%	DHR-EIS	1970-1976
402	LFN/POP	LABOR FORCE, RATIO OF EMPLOYMENT TO POPULATION			%	DHR-EIS	1961-1978
403	N	EMPLOYMENT, TOTAL BY PLACE OF RESIDENCE (INCLUDES PERSONS ON STRIKE)	132	121	TPPL	DHR-CPS	1970-1978
404	U	UNEMPLOYMENT, TOTAL (NOT SEASONALLY ADJUSTED)	133	122	TPPL	DHR-CPS	1961-1978
405	URATE	UNEMPLOYMENT RATE (NOT SEASONALLY ADJUSTED)	134	123	%	DHR-CPS	1961-1978
406	NRESADJ	EMPLOYMENT, RESIDENCE ADJUSTMENT FOR OUT-OF-STATE COMMUTERS	135	124	TPPL	DHR-EIS	1970-1978
407	NPW	EMPLOYMENT, TOTAL BY PLACE OF WORK (INCLUDES PERSONS ON STRIKE)	136	125	TPPL	DHR-EIS	1961-1978
408	NAG	EMPLOYMENT, AGRICULTURAL	137	126	TPPL	DHR-USDA	1961-1978
409	NAGH	EMPLOYMENT, NUMBER OF HIRED FARM WORKERS, KENTUCKY	138	127	TPPL	USDA	1951-1978
410	NAGNH	EMPLOYMENT, NUMBER OF NON-HIRED FARM WORKERS, KENTUCKY	139	128	TPPL	USDA	1951-1978
411	NNAG	EMPLOYMENT, TOTAL NON-AGRICULTURAL	140	129	TPPL	DHR-EIS	1961-1978
412	NNWSNAG	EMPLOYMENT, NON-WAGE AND SALARY, NON-AGRICULTURAL (SELF-EMPLOYED)	141	130	TPPL	DHR	1961-1978
413	NSTRKE	EMPLOYMENT, PERSONS INVOLVED IN LABOR-MANAGEMENT DISPUTES	142	131	TPPL	DHR-CPS	1964-1978
415	NNAGWS	EMPLOYMENT, TOTAL NON-AGRICULTURAL WAGE AND SALARY	145	135	TPPL	DHR-790	1951-1978
416	NMFG	EMPLOYMENT, TOTAL MANUFACTURING	146	136	TPPL	DHR-790	1951-1978
417	NMFG	EMPLOYMENT, MANUFACTURING, DURABLE	147	137	TPPL	DHR-790	1951-1978
418	NMND	EMPLOYMENT, MANUFACTURING, NON-DURABLE	148	138	TPPL	DHR-790	1951-1978
419	NNMNAGWS	EMPLOYMENT, NON-MANUFACTURING, NON-AGRICULTURAL WAGE AND SALARY	149	139	TPPL	DHR-790	1951-1978
421	NFOD	EMPLOYMENT, FOOD AND KINDRED PRODUCTS (SIC 20)	151	141	TPPL	DHR-790	1951-1978
422	NDLQ	EMPLOYMENT, DISTILLED LIQUORS, EXC. BRANDY (SIC 2085)	152	142	TPPL	DHR-790	1951-1978
423	NTOB	EMPLOYMENT, TOBACCO MANUFACTURES (SIC 21)	153	143	TPPL	DHR-790	1951-1978
424	NCIG	EMPLOYMENT, CIGARETTE MANUFACTURE (SIC 211)	154	144	TPPL	DHR-790	1951-1978
425	NTSR	EMPLOYMENT, TOBACCO STEMMING & REDRYING (SIC 214)	155	145	TPPL	DHR-790	1951-1978
426	NTEX	EMPLOYMENT, TEXTILE MILL PRODUCTS (SIC 22)	156	146	TPPL	DHR-790	1951-1978

TABLE 2. Relationship of Personal Income and Taxable Income—United States (billions of dollars).

		1974
Personal Income (BEA)		1153.3
Less:	Portion of personal income not included in adjusted gross income	273.5
	Transfer payments (except taxable pensions and military retirement pay)	127.5
	Other labor income (except fees)	54.6
	Imputed income	51.3
	Other types of personal income	40.0
Plus:	Portion of adjusted gross income not included in personal income	81.6
	Personal contributions for social insurance	47.6
	Net gain from sale of capital assets	13.2
	Capital consumption adjustment (except on owner-occupied dwellings)	4.6
	Other types of income	16.3
Equals:	Adjusted gross income of all individuals, estimated from personal income (BEA)	961.5
Less:	Difference between BEA and IRS estimates of adjusted gross income	55.4
Equals:	Adjusted gross income reported on individual returns (IRS)[a]	906.1
Less:	Adjusted gross income reported on nontaxable returns (IRS)	25.9
Equals:	Adjusted gross income reported on taxable returns (IRS)[a]	880.2
Less:	Deductions on taxable returns (IRS)	171.4
	Standard Deductions	57.5
	Itemized Deductions	113.9
Less:	Exemptions on taxable returns (IRS)	136.6
Plus:	Tax preferences adjustment[b]	0.3
Equals:	Taxable income on taxable returns (IRS)	572.4

[a]Net of deficits.

[b]Beginning in 1970, with the establishment of the tax preferences, individuals subject to this tax are classified as taxable individuals even if they have negative taxable income (Adjusted Gross Income less deductions and exemptions). Such individuals are included in the IRS *Statistics of Income* tabulations as having taxable incomes of zero, rather than the negative amount. Therefore, taxable income of taxable individuals shown in *Statistics of Income* is greater than the amount obtained by subtracting deductions and exemptions from Adjusted Gross Income. The entries on this line show the amount of the difference.

SOURCE: *Survey of Current Business.* 56(12):18; December 1976.

tion of series can occur in practice, but because of the institutional characteristics of data provision by the original sources; economic data are often subject to revison and, sometimes, observations differing only in the revision cycle represented may need to be stored simultaneously [19].

Expansion of the database is also evidently facilitated by its subdivision into data banks—although, it should be noted, physical modularity of the database does not require logical modularity. A unitary database can span physical data sets and, similarly, it would be possible to use one physical data set to hold a number of data banks. Nevertheless, because the KEIS public database is organized into data banks which individually span at least two physical data sets, expansion can occur without there being any technical limit on the extent of the database; the hardware support for the KEIS includes an IBM 3850 Mass Storage Subsystem for the quasi-offline storage of seldom used banks. Furthermore, the combination of logical and physical modularity tends to minimize the effects of database growth on the use of previously included data series.

Finally, it is clear that "comprehensibility" is aided by the partitioning of the database: not only are smaller, more specialized banks individually more easily comprehended than a single, larger, composite database, but partitioning can reveal to users the common characteristics of groups of data, thus making the database as a whole more understandable.

However, there are also specific comprehension-aiding benefits of the partitioning. One of these, for example, is that as a consequence it may be possible to simplify the naming of data series: the storage in separate data banks of series referring to the same geographic level, but different frequencies of observation, allows names to be made common across frequencies, yet to be unique within a given bank. As a result, series can be mnemonically named in conformity with a standardized naming convention, something much more difficult if not impossible to achieve if each series had to be given a distinct name consisting of a small number of alphanumeric characters.

It is important for a user to be able to remember names easily, for these can appear in the KEIS data retrieval commands in the context of complex mathematical expressions, as well as on their own. But, in addition, apart from any database considerations, it can be argued that data series that differ only in frequency of observation should carry the same name, for they represent measurements on the same economic concept. This argument is further buttressed by the fact that there are very few instances in which it is appropriate to mix data of different frequencies. Consequently, although the KEIS software allows data of different frequencies to be retrieved from different banks, transformed to a common frequency, and then used together, it generally does not permit the simultaneous processing of data that do not have a common frequency. The important exception to this rule is in the case of the printing of certain tables.

The Provision of Information About the Data: The User Manual

The way in which the KEIS programs allow a user to progress from one program level to the next was briefly described earlier. However, the fundamental problem a modern economic database system poses for the user who is not an economist is not discovering how to operate it mechanically but learning how to use economic data. Once a person begins to execute programs that allow the discretionary retrieval of data, he must know specifically which data he wants to retrieve. In any particular case, the data obtained could be one or more observations withdrawn from the database or else the result of some type of mathematical operation; if the latter, the user will generally need to specify both the particular transformation to be made and the one or more data items that are to be transformed.

TABLE 3. Comparison of Kentucky coal mining employment data from three different sources for three years.

Year	Source	Underground Mining	Surface Mining	Strip Mining	Auger Mining	Auger-Strip Mining	Total
1973	KDMM[a]	20,159	10,402	4411	488	5503	30,561
	BOM[b]	19,584	8391	6913	1478	d	27,975
	BLS[c]	d	d	d	d	d	26,500
1974	KDMM	24,279	15,014	5817	580	8617	39,293
	BOM	22,575	14,470	d		d	37,045
	BLS	d	d	d	d	d	34,000
1975	KDMM	26,793	21,612	8381	480	12,751	48,405
	BOM	22,200	13,870	d		d	36,070
	BLS	d	d	d	d	d	41,500

[a]Kentucky Department of Mines and Minerals.
[b]U.S. Bureau of Mines.
[c]Bureau of Labor Statistics, 790 Monthly Survey of Establishments.
[d]Data collected under this program are not disaggregated to this category.

Ideally, any person using an economic database system should know the precise characteristics of any data that he retrieves, but at the very least he must know what the data are called by name and/or where they are in the database, if he wants to use the system features to any appreciable degree.

Ostensibly, the problem is one of data searching. However, finding a particular series can be accomplished in a variety of ways, including using a data bank index or, more rapidly, by combining some form of keyword search by broad data categories, such as income, prices, and the like, with a local index search. For at least some users, the much more fundamental problem is not how to find data, but knowing specifically what to look for within a data category. The nature of this problem is easily demonstrated: Table 2 shows aggregate U.S. income for 1974 as measured by the U.S. Department of Commerce and the Internal Revenue Service of the U.S. Department of the Treasury. This table also reveals the definitional differences between the measurements. Note, for instance, that Personal Income does not include capital gains but that Adjusted Gross Income does. Note also that there are two IRS concepts of income—Adjusted Gross Income and Taxable Income—and that Taxable Income, in particular, is in magnitude about half of Personal Income. Clearly, it does matter which concept of income is used and, in any particular case, a person using an economic database system containing all three measurements must know or be told which is the relevant concept of income to choose.

In addition to being associated with a particular conceptual definition, economic data have a number of other characteristics that tend to make these data difficult for the noneconomist to use effectively. A fundamental characteristic is that certain types of data are scarce. This scarcity is partly a result of the circumstance that government agencies are the original source of much of the available economic data and that although some agencies collect data for the specific purpose of monitoring economic activity, most collect only those data that are necessary for their own administrative needs. However, even if data are collected, this does not imply their publication. In addition to government agencies, private individuals, firms, and other economic entities all collect data. For any one of several reasons, most of the economic data collected are never published; the reasons include confidentiality restrictions and the costs of compiling and publishing tables of data. One of the most important consequences of data scarcity is that frequently the data people seek are not available, at least as reliable estimates. One of the most important consequences of the dependence on administrative records as a basic economic data source is that many important summary statistics, including Gross National Product, are not necessarily constructed from the most appropriate component measurements; this affects the reliability of the estimates made.

In contrast, there are certain data that are independently collected and reported by several original sources. For example, in Kentucky there are three independent sources of data on state coal mining employment and, as shown in Table 3, the differences between the estimates can be substantial. This multiplicity in the publication of particular data raises the issue, in the development of an economic database, whether to select one source for each type of data. For many users of economic data, discriminating between sources is difficult, especially in those instances in which competing data series are ostensibly measurements on the same economic concept. Actually, there is no easy solution to the problem, for including only data from a particular source is a form of censorship; carrying data from all competing sources implies the need for a substantial user education effort. Of course, ideally, the issue should be the accuracy of the competing estimates; however, not only is it usually the case that there are conceptual differences involved, but measuring the relative accuracy of the alternative estimates may not be possible.

It is important to realize that economic data are seldom "evaluated" data. "Evaluated" data are those that are the result of repeated, independent measurements that are com-

bined appropriately to produce a statistic; for example, measuring the length of a room repeatedly in various, appropriate ways, then taking the mean of the individual measurements produces an "evaluated" estimate of the length of the room. Characteristically, aggregate economic measurements can be made only once under a given set of conditions. Consequently, as a rule, it is not possible to assign a numeric assessment of the reliability of the data to a particular aggregate economic data series. Under such conditions, any assessment of reliability must be made in terms of the reliability of the methodology used, not the accuracy of the particular estimates.

Finally, as noted earlier, it is also common for economic data to be subject to revision as new, and hopefully better, estimates become available. Often, preliminary reports are made before the final estimates are obtained, either as the result of an attempt to be timely in reporting or because of a significant delay in the acquisition of component data. Occasionally, revisions can be made because of an administrative change in the conceptual definition or the classification of particular data items.

Only some of the characteristics of economic data have been considered, but it should be clear that any user of economic data should know the concepts involved in any data he uses and that he also should understand the particular measurement methodologies used and their reliability. Unfortunately, this information can be difficult to obtain. In certain instances, as in the case of state and county Personal Income estimates or, alternatively, estimates of employment, there is a federal government agency responsible for the production of descriptive information concerning the data most commonly used. However, while a literature exists that provides adequate information about the conceptual characteristics of the data [20-23], it generally requires some degree of expertise to ferret out detailed descriptions of the current estimation methodology [24, Part II]. It can require considerable expertise to determine the reliability of that methodology [24, Secs. 2.3 and 2.4; 25,26]. In the case of other data, the problems of obtaining information can become considerably more severe, particularly for those who are not experienced users of economic data; even for those who are expert, obtaining such information can be a long and tedious process.

Because of the information needs of users of the KEIS and the difficulties most would have in obtaining that information independently, the KEIS *User Manual* [24] provides considerable background information about the data the public database contains. Part II of this manual, which describes the data, presently runs to several hundred pages and is the largest of the manual's four parts. It explains the concepts involved, the measurement methodologies used by original sources of the data, and provides assessments of those methodologies. However, not every series in the public banks is described; only those that have some unique characteristic. The focus is upon categories of data and the stress is upon those data for which no alternative description is currently available in published form. The purpose is not to repeat detailed information that is otherwise easily obtained. Instead, the intention is to document the gaps, to provide references as appropriate and, in general, to make the reader of the manual aware of the particular characteristics of economic data as a class.

Obviously, providing this type of service to users is a long-term project, which by its nature is never completed. Today, the *User Manual* contains a description of the major federal and federal-state data collection programs, but only some of the sections describing the construction of particular categories of data have been issued. Providing up-to-date documentation is always difficult, but in this case there is the additional problem that providing the data descriptions usually involves extended discussions with the original data sources in order to determine the precise constructive characteristics of the data. Furthermore, before they are issued, these sections are reviewed by independent experts, as well as by the original data sources, itself a time-consuming process.

Using the KEIS

The KEIS programs are operated using a command language that is designed to provide both power and flexibility. *Power* refers to the degree of integrated processing that the software allows. *Flexibility* refers to the degree of control that the user has over the particular operations performed. The power of the KEIS software can be illustrated by considering a command of the form $Z = X/Y$. A command of this type causes both the retrieval of X and Y from a data bank and the division of X by Y to obtain Z. Normally, the results Z will be displayed, but, under certain conditions, this command will also cause these results to be stored, either in the same data bank from which X and Y were retrieved or in another bank. The flexibility of the command language is indicated by the fact that the user alternatively can cause these operations to be carried out as a sequence of separate steps. The user determines the data banks from which the data are retrieved, as well as the private bank in which the results might be stored. He determines whether X, Y, and Z, as symbolic names, reference scalars or data vectors. He specifies the particular operations to be performed with the data.

This simple example does not define the limits of the power of the KEIS software. For instance, it is also possible, when using an econometric model, to estimate or reestimate a particular equation and then directly embed it into the model so as to allow the modified model to be used immediately. Alternatively, such an operation can be carried out as a sequence of separate processing steps.

Such wizardry should not, however, be allowed to obscure the fact that the user of the KEIS must first identify the data to be used. Data retrieval is inevitably a two-step procedure, even if the physical act of retrieval can be made an integral part of the data-processing step. Furthermore, in the case of an economic database, searching for particular data does have the characteristic that to find data easily almost requires knowing where they are beforehand, in terms of either physical location or symbolic position.

TABLE 4. State data banks data series organization chart.

Data Group	Annual Bank	Quarterly Bank	Monthly Bank
Personal Income Data (includes personal income, earnings data)	1–399	1–199	1–124
Manpower Data (includes all employment, unemployment, labor force, and hours data)	400–590	200–424	125–424
Population and Other Resources Data	591–599	425–449	
Final Demand Data (includes macroeconomic estimates for the state, e.g., gross state product)	600–699	450–524	
Manufacturing Industry Data (SIC 20-39)	700–999	525–549	
Mineral Industry Data (SIC 10-14)	1000–1224	550–749	425–474
Contract Construction Industry Data (SIC 15-17)	1225–1374	750–799	475–499
Wholesale, Retail Trade, and Service Industry Data (SIC 50-59, 70-89)	1375–1400	800–824	
Transportation, Communication, Public Utilities Industry Data (SIC 40-49)	1401–1479	825–924	500–649
Finance, Insurance, Real Estate Industry Data (SIC 60-67)	1480–1499	925–1324	650–674
Agriculture Data	1500–1999	1325–1424	675–749
Miscellaneous, Run-Over Data	2000–2499	1425–1524	750–849
By-Product of Government Activity Data	2500–2999	1525–1549	850–874
Government Revenue and Expenditure Data	3000–5799	1550–1899	875–1174
National Data	5800–6000	1900–2000	1175–1200

Searching the Database

Given that the user can identify the relevant data bank, there are two methods of searching the KEIS public database. The first involves the use of an organization chart that defines the relative locations of categories of data. The chart for the state banks has been reproduced from the *User Manual* as Table 4. Note that this table identifies the categories in these banks within series number groupings and that, for example, the annual data on Personal Income and its components are to be found within the range of series numbers 1–399 in the Annual Bank. The index to the Annual Bank obviously can be used to find any particular series in this category. Subcategories of data, such as Wage and Salary Income, Labor and Proprietors Income, or Transfer Payments, are grouped together in the index, so that even a relatively inexperienced user of the KEIS can find a particular series quickly.

Alternatively, the data bank indices can be computer-produced by the user alphabetized by series name. As a result it is possible to search for series by name, especially since the names are determined by a mnemonic convention. For example, series names that begin with the letter N are observations on some form of employment; the remaining characters of the name generally define the particular employing industry. Similarly, names that begin XP are observations on physical production. Because the series names are constructed in this way, an alphabetic index can be as useful a searching tool as a sequential index; and not only can the user determine from the naming convention what a name should be, then look for it in the alphabetical index, but such an index provides an alternative category ordering of the data series in a particular bank. For instance, an alphabetical index will show all measurements on physical production as a group, irrespective of industry; although, within this group, there will still be an industry ordering, alphabetized by the industry codes which form the remainder of the series names.

There are also a number of other searching and information aids for users. One of these is the partial index. Sometimes, a user will wish to review all the data in a given bank that are obtained from one (or several) original sources.

Note the column of source codes, headed SOURCE, in the Annual Bank index portion in Table 1. These, together with the UNITS codes next to them, are defined on two initial pages that are automatically computer-printed with each of the indexes; these codes are also reproduced in the *User Manual.* Using the source codes, it is possible to produce an index of only those data in the bank that are obtained from the specified sources.

Similarly, the facility exists to produce only a part of the County Bank index. Often users of the County Bank will wish to use just the data pertaining to a particular county or certain aggregates of counties. Since printing a partial index may cost only a few cents, rather than $10 or more for the complete index, most users of the County Bank work from partial indexes.

The data bank indexes each show the last date of any change in the bank. However, a user may wish to know the last time a particular data series was either revised or extended. Another system feature is the *Data Bank Log.* This can be printed by a user at any time for any bank and gives the last dates of revision and extension of each of the data series. For a given bank, the user controls how much of the log is printed. Incidentially, this information also appears in other contexts, as will be illustrated.

The focus has been upon the public banks. However, the facilities that have been described are generally available in the case of private banks as well. The creator of a private bank is responsible for naming the series in the bank, as well as describing them and defining their sources and units. Consequently, the degree to which an alphabetic and a sequential index for a given private bank will be equally useful searching devices will depend on the creator of the bank. Similarly, consistency in defining the source codes for the data will determine whether a particular partial index will be useful. In contrast, the computer automatically keeps track of data series revisions and extensions for private banks as well as public banks.

A Simple Example of Data Retrieval and Manipulation

As has been described, the KEIS offers considerable scope for the use of economic data in various ways, some of which uses are quite sophisticated. However, for many users, the most common operation performed is the retrieval of one or more observations on some economic variable or the algebraic manipulation of two or more data series. These are also operations that are easily illustrated.

The particular commands to execute a program depend on the computer network and are described in Parts III and IV of the *User Manual.* However, they may be as simple as the command "EXECUTE," followed by the program name. In order to retrieve a series of observations on Kentucky Personal Income from the Annual Bank for the period from 1958 to 1977, the full set of commands might be:

```
EXECUTE DTP
INPUT BANK=ABANK
SET DATES=1958-1977
QUERY: Y$(1)
END
```

To also compute per capita personal income and store the results in a user's private bank, they might be:

```
EXECUTE DTP
INPUT BANK=ABANK
OUTPUT BANK=USERBANK
SET DATES=1958-1977
QUERY: Y$(1)
STORE RESULTS
YPC$(3)= Y$(1)/POPCIV(591)
END
```

The commands are almost self-explanatory once it is realized that ABANK is the code name of the Annual Bank, Y$ and POPCIV are the series names for Personal Income and the civilian population, and the series numbers for each of these in the Annual Bank are 1 and 591, respectively.

The batch mode response to the second set of commands is shown in Table 5. For the most part, the relationship between the commands and the computer's response is obvious, but it should be noted that the Output Bank entry for YPC$—the series which results from the division of Personal Income by population—was created in a previous job. In order to establish a new data series entry in the Output Bank, as well as to carry out the operations illustrated, the extended set of commands might be:

```
EXECUTE DTP
INPUT BANK=ABANK
OUTPUT BANK=USERBANK
SET DATES=1958-1977
QUERY: Y$(1)
STORE RESULTS
DEFINE NEW SERIES: YPC$(3)
DESCRIP:INCOME, PER CAPITA PERSONAL
UNITS=$,SOURCE=SCB-T38,FREQUENCY=ANNUAL
YPC$(3)=Y$(1)/POPCIV(591)
END
```

These commands would generate printed output very similar in appearance to that shown in Table 5. The principal difference would consist of an additional message to the effect that a new series, YPC$, had been defined as series number 3 in the Output Bank. This message would appear just after the line "RESULTS THAT ARE STORED ARE NOTED BELOW."

The several examples just given provide some indication of the syntactical characteristics of the KEIS command language, but an additional comment might be made concerning the mechanics of the retrieval and storage process for data. It will be noted that the series name and number are both required, as mentioned earlier. As an alternative, it would have been possible to have allowed the series name alone to be used; in fact, it is common for other economic database systems to reference data in this way. However, while requiring the user to include both name and series number each time obviously imposes a greater burden on him than requiring the name only, the latter increases the likelihood of a data retrieval error as the result of the misidentification of series, particularly if series names are similar, which they frequently are. In contrast, the joint use

TABLE 5. Kentucky Economic Information System: DTP—Data Transformation and Retrieval Program
15:59:38 Tuesday, 20 Nov 1979.

```
CHARACTERISTICS OF INPUT BANK:
      BANK DESCRIPTION: KENTUCKY ANNUAL ECONOMIC DATA BANK
      NUMBER OF SERIES IN BANK: 4044
      BANK LAST UPDATED AT 11:56:50 on 20 NOV 1979
CHARACTERISTICS OF OUTPUT BANK:
      BANK DESCRIPTION: USER'S PRIVATE DATA BANK
      NUMBER OF SERIES IN BANK: 58
      BANK LAST UPDATED AT 10:17:18 ON 20 NOV 1979
DATES SET. PERIOD IS: 195801-197701
OPERATION PERFORMED:
            QUERY: Y$(1)
     SERIES: Y$      SERIES NUMBER:     1      UNITS: M$      SOURCE: SCB-T38
INCOME, TOTAL KENTUCKY PERSONAL, BY PLACE OF RESIDENCE
                                ANNUAL DATA
1958-1963    4412.137    4641.617    4793,527    5075,426    5374.512    5645.922
1964-1969    5917.758    6427.531    7041.848    7621.426    8358.996    9170.172
1970-1975    9937.332   10744.453   11891.098   13395.895   15174.211   16537.063
1976-1977   18535.547   20656.012
LAST REVISED: 11/08/79
LAST EXTENDED: 10/04/79
THE ABOVE RESULTS HAVE BEEN PRINTED ONLY
****
STORE RESULTS COMMAND RECEIVED.
      RESULTS THAT ARE STORED ARE NOTED BELOW

OPERATION PERFORMED:
          YPC$(3)=Y$(1)/POPCIV(591)
RESULT OF TRANSFORMATION
                                ANNUAL DATA
1958-1963    1511.006    1568.114    1598.909    1685.069    1773.181    1848.093
1964-1969    1918.858    2074.736    2270.100    2444.332    2664.647    2912.091
1970-1975    3121.022    3320.289    3638.647    4071.700    4576.059    4927.609
1976-1977    5454.840    6020.402
ABOVE RESULTS STORED IN OUTPUT BANK AS SERIES NUMBER:  3, CALLED: YPC$.
SERIES DESCRIPTION:
      INCOME, PER CAPITA PERSONAL
  UNITS: $
  SOURCE: SCB-T39
AS STORED, SERIES AVAILABLE FOR PERIOD 195101 TO 197701
****
END OF JOB -- SUMMARY OF OPERATIONS:
      8 CARDS READ.
      1 QUERIES MADE
      1 ARITHMETIC OPERATIONS PERFORMED
STATUS OF OUTPUT BANK AT THIS TIME:
      BANK TITLE: USER'S PRIVATE DATA BANK
      EARLIEST OBSERVATION IN BANK: 195101
      MAXIMUM NUMBER OF OBSERVATIONS POSSIBLE ON ANY SERIES: 400
      MAXIMUM NUMBER OF SERIES POSSIBLE: 1200
      NUMBER OF SERIES NOW IN BANK: 58
      LARGEST SERIES NUMBER: 72
      BANK LAST UPDATED AT 15:59:38 ON 20 NOV 1979
JOB COST      $.11      13 EXCPS      223 LINES SPOOLED      .00011 HOURS CPU TIME
```

of name and number allows cross-checking and tends to insure the immediate discovery of series identification mistakes.

Preventing such errors is especially important at each end of the use spectrum. Users of an economic database system will not necessarily be able to recognize immediately that such an error has been made: the plausibility of the numbers obtained may be the only indication of the error and the less sophisticated the user the less likely he is to be able to judge this. However, in the construction of an

econometric model, it is possible to use inappropriate data in the estimation of one or more equations and for the effects of this to be not immediately apparent to anyone. For a model containing variables that are intricate transformations of other variables, it may not be until the model is put together and tested against other, appropriate data that any errors made become obvious. At this point, particularly if the model is large, it may even take days to discover the source of an error, for errors can be made at any stage of the model construction process.

Conclusion

Various aspects of the Kentucky Economic Information System have been described, but the system features that have been emphasized are those that make the KEIS usable by a variety of people. This focus of attention is appropriate for at least two reasons: the first is that there is a growing general interest in economic data. In the United States, interest in these data has been stimulated by legislation and regulations that cause government grants-in-aid and other disbursements to be distributed on the basis of the values of certain regional economic statistics and require the filing of economic impact statements by many organizations. But even apart from such induced interest, there is now in every industrial country a heightened consciousness of economic data and of the performance of national and regional economies. Furthermore, both governments and businesses are increasingly likely to want to use economic data in their forecasting and planning. Consequently, the increase that occurs in the next few years in the number of economic database systems is likely to reflect not simply the demands of economists but also the requirements of others.

The second reason this focus of attention is appropriate is that developing an economic database system for use by a variety of people represents a much greater challenge than developing a system for the use of economists alone. As has been described, to use economic data effectively generally requires knowledge of the particular characteristics of these data and of the ways in which the data should be presented and used. Therefore, it is not enough simply to make an economic database system easy to use, in terms of the mechanical operations involved. There is a need to include system features that will allow even the least sophisticated user to obtain output that he finds comprehensible. The question is, to what extent is this a reasonable objective? It is obvious that there is a limit to the degree to which a machine can be substituted for human training and intelligence: economic and other database systems that contain specialized data, the use of which may involve specialized techniques, cannot necessarily be developed to the point that they can meet the particular needs of anyone who might wish to use them. Nevertheless, as in the case of the KEIS, it may be possible to develop systems that are sufficiently broad in their capabilities to be usable to some extent by many people.

The KEIS has been described as an economic database system. This characterization refers to the data it contains and the analytic techniques it makes available to users. However, while the system is intended to be used with time series data, it is not restricted necessarily to use with economic data. Broadening the content of the database slightly and adding more analytic features would allow the KEIS to be considered in a more general context, as a socio-economic database system or, possibly, a social science database system. That types of numeric database systems are not always sharply distinguishable from each other needs to be recognized, if only to clarify that the essential characteristic of a particular system can be how the data it contains happen to be interpreted and used.

References

1. Luedke, J. A.; Kovacs, G. J.; Fried, J. B. "Numeric Data Bases and Systems." *Annual Review of Information Science and Technology*. White Plains, NY: Knowledge Industry Publications; 1977: 12:118-181.
2. Pilozzi, G. "A New Approach to the Dissemination of CANSIM Data." *Canadian Statistical Review*. 51:5-8; June 1976.
3. Triandafillou, P. N. "CANSIM/SIS—A New Phase of Development." *Canadian Statistical Review*. 50:4-6; January 1975.
4. Darrow, J. W.; Belilove, J. R. "The Growth of Databank Sharing." *Harvard Business Review*. 180-190; November-December 1978.
5. Doebler, P. "Data Base Publishers Vying for Key Roles as Rapid Growth Looms in Business Uses." *Publishers Weekly*. 38-40; 16 September 1974.
6. King, H. "A Review of State and Local Government Census Data Processing Software Needs." *Review of Public Data Use*. 7(1):2-14; January 1979.
7. National Association for State Information Systems. *Information Systems Technology in State Government*. Lexington, KY; 1979.
8. Urban and Regional Informations Systems Association. *Annual Conference Proceedings*. An annual publication.
9. Intriligator, M. D. *Econometric Models, Techniques, and Applications*. Englewood Cliffs, NJ: Prentice-Hall; 1978.
10. Renfro, C. G. "The Kentucky Quarterly Economic Model: Recent Developments." In Mickle, M. H.; Vogt, W. G., Eds. *Proceedings of the Eleventh Annual Pittsburgh Conference on Modeling and Simulation*. Pittsburgh, PA: Instrument Society of America for the University of Pittsburgh; 1980, in press.
11. Renfro, C. G. "Economic Data Base Systems: Some Reflections on the State of the Art." *Review of Public Data Use*. 8(2); August 1980.
12. Lynch, L. K. "The Long-term Growth of the Kentucky Economy: 1979 Update." *Kentucky Council of Economic Advisors Studies in Applied Economics, Number 1*. Lexington, KY; January 1980.
13. Kentucky Council of Economic Advisors. *Annual Report*. Frankfort, KY: Office of the Governor. An annual publication.
14. Hultman, C.; Ramsey, J. *Kentucky Energy Consumption and Requirements*. Lexington, KY: Energy and Critical Resources Center, University of Kentucky; 1975.
15. *Kentucky Economy: Review and Perspective*. Lexington, KY: Center for Applied Economic Research, University of Kentucky, for the Kentucky Council of Economic Advisors. A quarterly publication.
16. Kentucky Department of Commerce. *Kentucky Deskbook of Economic Statistics*. Frankfort, KY. An annual publication.
17. Chapman, C. A. "Situation Report: the Development Information System for Kentucky." In: Chapman, C. A.; Infanger, C. L.; Robbins, L. W.; Debertin, D. L., Eds. *Taking Computers to the Community: Prospects and Perspectives*. Lexington, KY: University of Kentucky College of Agriculture Cooperative Extension Service; 1978: 187-190.

18. Maki, W. R.; Turnquist, R. E.; Venegas, E. C. "Minnesota Energy-Economic Information System." Paper prepared for the Invited Paper Session on State Energy Modelling, ORSA/TIMS 1976 Joint National Meeting, Philadelphia, PA, March 31–April 1, 1976.
19. Renfro, C. G. "On the Development of a Comprehensive Public Data Base for Aggregate State and Local Economic Data." *Review of Public Data Use.* 7(6); December 1979.
20. Bureau of Labor Statistics. *Employment and Earnings.* Washington, D.C.: U.S. Government Printing Office. A monthly publication.
21. Bureau of Labor Statistics. *Handbook of Methods.* Bulletin 1910. Washington, D.C.: U.S. Government Printing Office; 1976.
22. Goldstein, H. *State and Local Labor Force Statistics.* Washington, D.C.: National Commission on Employment and Unemployment Statistics; May 1978.
23. Schwartz, C. F.; Graham, R. E. *Personal Income by States Since 1929, A Supplement to the Survey of Current Business.* Washington, D.C.: U.S. Government Printing Office; 1956.
24. Renfro, C. G.; Coomes, P. A. *Kentucky Economic Information System User Manual.* Lexington, KY: Office of Research, College of Business and Economics, University of Kentucky; 1977. (Cost: $20 postpaid, Manager, Kentucky Economic Information System, Center for Applied Economic Research, College of Business and Economics, University of Kentucky, Lexington, KY 40506.)
25. Lynch, L. K. "Review of BEA Methodology for Personal Income Estimates." *Kentucky Council of Economic Advisors Policy Papers Series, Number 1.* Lexington, KY: Office of Research, College of Business and Economics, University of Kentucky; February 1977.
26. Reid, J. N. "Understanding Federal Programs: The Need for a Coordinated Data System." *State and Local Government Review.* 11(2):42–47; May 1979.