

The California State University

**SOCIAL SCIENCES RESEARCH
and
INSTRUCTIONAL COUNCIL**

**18th ANNUAL REPORT
1989 - 1990**

The 1989 - 1990 academic year was an active and productive one for the SSRIC. Several significant ongoing projects were continued and new ones initiated. The Council held three business meetings during the year: The Fall 1989 at CSU-Hayward, Winter 1990 at CSU Fullerton, and culminating in a highly successful student research conference at Cal Poly Pomona in May 1990. In-house workshops were held at the Hayward meeting (on UNIX) and at the Fullerton meeting (on the VAX). Scheduled meetings for the 1990 - 1991 years are October at CSU Fresno; February, at CSU-Dominguez Hills and in May at CSU - Hayward. Ted Anagnoson, LA State was selected Chair for 1990-91 and the Council also welcomed Frank Gosset as the representative from CSU-Long Beach.

FIELD REPORT:

Each year the SSRIC selects and recommends faculty and students for internships at the Field Institute and evaluates proposals from faculty who wish to include some questions in one of the California Field polls. This year the SSRIC Field Committee consisted of Ed Nelson (Fresno), Bruce Haston (Humboldt) and Richard Serpe (Fullerton). The student intern selected was Mary McCullen (SLO) who served through the summer of 1990. A faculty fellowship was awarded to Professor Elizabeth Nelson (Fresno) for her project: "The Changing Role of Women." The Committee also recommended that six question credits be awarded to Joel Best (Fresno) for his research on Satanism.

During the year the Institute planned five statewide surveys. Election contests as well as a host of other political topics were covered. It was noted that the delivery of released surveys and the general management of the surveys now seems to be smoother than previously. There was, however, concern expressed about Field's proposal to use a split form questionnaire

during election year surveys. Field remained committed to splitting the political questions from the social questions because of the increased length of the interview schedule. Nevertheless he did promise to include some political questions on the general social question survey.

ICPSR: ANN ARBOR MEETING

The biennial meeting of the ICPSR was held at Ann Arbor in November of 1989. Attending were Dan Graves, Chair (Hayward), Jeff Johnson (CCR), John Korey (Pomona), Richard Serpe (Fullerton), and Elliot Barkan (San Bernardino). A serious matter of discussion at the meeting was the impending split between the ICPSR and ISR. The SSRIC expressed its concern over the potential negative effects of such a separation, especially if it caused some key ICPSR personnel to resign. Carolyn Geda, Director of Management and Administration, discussed the problem with the Council at our Spring meeting. She explained that the ICPSR wanted more autonomy. The Council also discussed various strategies for increasing SSRIC influence at ICPSR. One suggestion was to chose a core of people who would always attend the ICPSR meetings, thus maintaining continuity. Another complementary solution would be to attempt to elect one of our members to the ICPSR Council. In related ICPSR matters, it was noted that the ICPSR was planning to change the way it distributes data and codebooks. Most notable was the addition of CD-ROM based data and codebooks. ICPSR also announced that it was willing to distribute the SSIMS modules once the authors agreed upon a format for that distribution.

ICPSR: SUMMER PROGRAM

Jon Ebeling (Chico) administered the ICPSR summer program. It was recommended and approved that the SSRIC support two candidates at \$1050 each. Those awards were made to Peter Haas (San Jose) and Bruce Wallin (Fullerton).

ICPS: WEST COAST SUMMER WORKSHOP

In a very exciting move, the SSRIC and ICPSR initiated negotiations to hold a jointly sponsored summer workshop on the West Coast which would closely parallel the annual ICPSR summer program held in Ann Arbor. SSRIC members attending the November 1989 OR meeting discussed the idea with Henry Heitowit, Director of the ICPSR Summer Program. Richard Serpe (Fullerton) was designated the SSRIC contact/negotiator. A memo was sent to Heitowit proposing a pilot session to be held summer 1991. A subcommittee was appointed consisting of Korey, Serpe, Dunn, Johnson, Ross, Barkan and Graves. Discussions continue.

CSU-FACULTY QUESTIONNAIRE:

Ted Anagnoson (Los Angeles) with the assistance of other Council members developed a questionnaire for a survey of social science faculty use of computing. The questionnaires were mailed to every chair of a Social Science department in the CSU system as well as to those on a SSRIC list of faculty known to be interested in social science computing. A preliminary report was issued in May, 1990 and a full report will be forthcoming during the 1990/91 academic year.

SSRIC FACULTY WORKSHOPS:

During the year the Council sponsored several successful and productive workshops for CSU faculty:

1. Ed Nelson reported on the workshop he presented in Fresno during June of 1989. There were 27 participants in two tracks: 12 in the beginners and 15 in the advanced. Those faculty in the advanced track worked with SPSS-PC and learned how to develop their own instructional modules for classroom use. Those in the beginning track worked with SPSS-PC and the SSIMS modules. Assisting Ed Nelson was John Korey (Pomona), Richard Serpe (Fullerton) and Jeff Johnson (ASG).

Ed Nelson also submitted a proposal for a SSRIC computing workshop to be held on the Fresno campus from June 24 through June 28, 1991. Over 60 faculty throughout the system indicated an interest in attending this workshop.

2. Ed and Elizabeth Nelson (Fresno) received a dissemination grant from the Chancellor's office to hold workshops to teach faculty how to use the computer modules developed by the SSIMS group. Workshops were held at San Francisco State University, CSU Fullerton, CSU Stanislaus and San Jose State University during the Spring of 1990.
3. A highly successful intermediate level workshop on Exploratory Data Analysis (EDA) was held at Fullerton January 26 through January 28, 1990. Organized by Richard Serpe (Fullerton) and Rich DeLeon (San Francisco State), this workshop was aimed at faculty who teach statistics. There were 43 applications and 23 who were accepted. The workshop was evaluated very highly by all participants.

SPRING STUDENT RESEARCH CONFERENCE:

Each year the SSRIC sponsors a student research conference in which both graduate and undergraduates present their papers in a convention setting. This year's first rate and highly professional conference was organized by John Korey and held at Cal Poly Pomona on May 10, 1990. This was the 15th

annual SSRIC student conference. Twenty-five papers were submitted and presented in six different panels. The keynote address, "Social Science in Action: Where It's Used and How", was given by Sandra H. Berry, Director Survey Research Group, the Rand Corporation. Awards of \$100 each and plaques were give in the following categories:

THE BETTY NESVOLD AWARD FOR BEST GRADUATE PAPER

Zack Loukides (CSU Fullerton)

"The Components and Correlates of Political Sophistication"

Jennifer Barnes (CSU Bakersfield)

"The Migration Myth: A Comprehensive Analysis of the Emigrations of the Nineteenth Century Irish and the Twentieth Century Oklahomans"

CHARLES McCALL AWARD FOR BEST UNDERGRADUATE PAPER

Sandi Ray (CSU Bakersfield)

"An Analysis of Cheating at California State University, Bakersfield"

THE GLORIA RUMMELS AWARD FOR BEST USE OF CSU DATABASES

Nancy Bednar (CSU Dominguez Hills)

"Variables and their Effects on Voting Behavior"

CENTRAL SUPPORT:

The CCR staff continued to provide excellent data base management and technical support. As a testimony to this, at the Spring Student Research Conference Penny Crane (Office of Communication and Computing Resources) was honored for her long time devotion and support of the SSRIC Council and its activities.

A major innovation came this year with the installation of QIKSTAT on the Cyber 960 in the Spring of 1990. Prepared by ASG and CCR, this easy-to-use program acts as an interface between SPSSX and the social science databases of the systemwide Computing Center's Cyber 960. It is capable of generating statistics for thousands of variables; yet, no experience with stat programming is required.

In other CCR news, Penny Crane reported that CCR had been reorganized to reflect changes that had happened by default anyway. It was also in response to the new Central IBM 4381 which was dedicated to administrative computing. There now is a smaller academic computing section which will work on academic computing exclusively.

As with my many predecessors, I reiterate my thanks to all Council members for their support and cooperation. It is this spirit of cooperation and selfless dedication which has sustained the Council and contributed to its many significant achievements. It was my pleasure to serve as chair.

Respectfully submitted,

A handwritten signature in cursive script that reads "Daniel R. Graves". The signature is written in dark ink and is positioned above the typed name and title.

Dan Graves,
California State University, Hayward
Chair, 1989/90

**1989 - 1990
SSRIC MEMBERS
and CYBER E-Mail Numbers**

Bakersfield

Jim Ross HBLR001
Dept. of Sociology
CSU, Bakersfield, Ca 93306

Chico

Jon Ebeling EEFA002
Dept. of Political Science
CSU Chico
Chico, CA 95927-0455

Dominguez Hills

Margeret Blue
Dept. of Political Science
CSU, Dominguez Hills
Carson, CA 90747

Fresno

Edward Nelson FCLR003
Dept. of Sociology
CSU, Fresno
Fresno, Ca 93740

Fullerton

Richard Serpe LRSERPE
Social Science Res. Center
CSU, Fullerton
Fullerton, CA 92634

Hayward

Dan Graves BCLQOAC
Dept. of Political Science
CSU, Hayward
Hayward, CA 94542

Humboldt

Bruce Haston GFA0001
Dept. of Political Science
Humboldt St. University
Arcata, CA 95521

Long Beach

Frank Gosset
Dept. of Sociology
CSU, Long Beach
Long Beach, CA 90840

Los Angeles

Ted Anagnoson
Dept. of Political Science
CSU, Los Angeles
Los Angeles, CA 90032

Northridge

Gene Turner RCLP001
Dept. of Geography
CSU, Northridge
Northridge, CA 91330

Pomona

John Korey CGTP012
Dept. of Political Science
CSU, Pomona
Pomona, CA 91768

Sacramento

Carole Barnes NBLRAAA
Dept. of Sociology
CSU, Sacramento
Sacramento, CA 95819

San Bernardino

Elliot Barkan PFAAAAJ
Dept. of History
CSU, San Bernardino
San Bernardino, CA 92407

San Diego

Paul Strand
Social Science Res. Lab
San Diego State Univ.
San Diego, CA 92182

San Francisco

Rich DeLeon SBKUAAA
Dept. of Political Science
San Francisco State Univ.
San Francisco, CA 94132

San Jose

Barbara Norrander TFNAAEQ
Dept. of Political Science
San Jose State Univ.
San Jose, CA 95192

San Luis Obispo

Dick Shaffer DDAAADT
Dept. of Social Sciences
Ca. Polytechnic St. Univ.
San Luis Obispo, CA 93407

Sonoma

Don Dixon UFAAAAC
Dept. of Political Sciences
Sonoma State University
Rohnert Park, CA 94928

Stanislaus

Larry Giventer VCLA002
Dept. of Politics & Public Admin.
CSU, Stanislaus
Turlock, CA 95380

CCR Representatives

Jeff Johnson JJOHNSON
P.O. Box 3842
Seal Beach, CA 90740-7842

Jonah Mills JMILLS
P.O. Box 3842
Seal Beach, CA 90740-7842

Penny Crane PCRANE
P.O. Box 3842
Seal Beach, CA 90740-7842

Chancellor's Office

Terry Dunn
Analytic Studies
CSU, Chancellor's Office
400 Golden Shore
Long Beach, CA 90802

The Cost of the Centralized Social Science Data Archive and Computing Service at Information Resources and Technology (Chancellor's Office) vs. The Costs of Campus Based Social Scientific Database Services

By J. Theodore Anagnoson, Chair¹
 CSU Social Science Research and Instructional Council

January, 1991

I. Purpose.

The purpose of this report is to calculate what it will cost and what kinds of services need to be provided to support social science database usage on each individual CSU campus, as opposed to the centralized provision of those services by the Academic Computing Support Group (ACSG) at CSU's Chancellor's Office. I requested the cost information from the Chancellor's Office (IRT), but this report is my work rather than theirs.

One reason I am circulating this report is that there has been lots of cost information circulating among the Committee of the Whole group, and most of this information is based only on "hits" to databases or software. *That information underestimates -- in many cases, quite substantially -- what it will cost to provide social scientific services at a reasonable level for the instructional and research/professional activities of typical CSU faculty.*

I have developed cost information for two campus scenarios:

The first assumes that the campus is a normal social science data base using campus and requires a reasonable level of services. (Even at this level of services, however, access to Census and other data will be less available than at present through IRT.)

The second is for absolutely minimum service. Some additional supplies will be ordered, and faculty will have memberships in ICPSR and be able to acquire a small number of additional datasets, but overall, service will be markedly lower than at present.

II. Background.

IRT will be returning to the campuses an estimated \$260,000 associated with its social scientific database function. Each campus will receive a portion of this amount (the portion is based on student enrollment), with the average campus receiving \$13,000.

¹ Chair, Department of Political Science, California State University, Los Angeles, CA 90032-8226. 213-343-2230 or kblqaae@calstate.

The question we address here is: Suppose your campus receives its \$13,000 -- what faculty services will be provided for the \$13,000? How much more or less will it cost to provide a level of service comparable to what the Academic Computing Support Group provided before the decentralization?

III. Assumptions.

Assumption 1.

The biggest assumption here is that we are assuming your campus has a platform (i.e., mainframe, minicomputer, or workstation) on which it can place the databases that are currently available or will be ordered for faculty use in the future. That platform should be capable of processing some substantial files (e.g., Census data) or faculty on your campus will not have access to these files unless access is purchased from another source. More on the latter below. For the largest files, only mainframes of moderate size or larger are really practicable, although some campuses will have relatively clever staff who can process databases on smaller platforms.

One time costs not included in this report:

- Acquiring a "platform" capable of processing the larger databases.
- Acquiring sufficient disk space (or tape capacity) to hold the existing archives. These currently run 30 gigabytes and include over 2,000 files.
- Cataloging these files and making the list available to prospective faculty users.
- Converting the several utilities now available on the Central Cyber which are designed to facilitate access to these files. These utilities are discussed below.
- Preparing documentation for the catalog and utilities.

Campuses, of course, may decide not to convert the utilities and not to make available all of the existing data files. In fact, my own conclusion is that the amount of work required to do all this² is so large that most campuses, realistically, are not going to do all the conversions. The most realistic expectation is that campuses will move some small number of the data files from the Central Cyber archive (the files faculty say they are actively using), put out a list of these files to interested faculty, distribute one or more current Inter-University Consortium for Political and Social Research catalogs to departments or keep one catalog centrally at the computer center for faculty use³ and then order datasets as required.⁴

Thus, we are really here talking about ongoing costs exclusive of hardware.

² IRT cannot do this work, as the platforms which are going to be used on the various campuses are very different from campus to campus.

³ The catalog is over 1,000 pages long and is quite cumbersome to use, which is why IRT developed the computerized documentation files and utilities.

⁴ That assumes that the campus is a member of ICPSR each year (\$5,500 per year). Any year that the campus is not a member and does not pay its dues will necessitate the physical removal of all ICPSR datasets from computers on campus and no use -- research, teaching, other, etc. -- at all as the license for use each year depends on paying the \$5,500 dues.

Assumption 2.

Scope of the report. No business databases are included (e.g., Compustat, CRSP, etc.). Social scientific databases which are included are those from the Inter-University Consortium for Political and Social Research (literally thousands of files are available, ranging from government surveys to the investigations of individual faculty), Field Institute (chiefly Field Polls of the California public), the Roper Center for Public Opinion Research (which has some files not available from ICPSR).

The social scientific files, no matter what the source, generally have the same format and contain similar information (thus they constitute a data base in the traditional sense of the word's meaning). They can be placed on the same platform at the individual campus level.

IV. A Campus Social Science Data Base Library.

All of the files put together constitute a data base library of social science information. The ideal computing platform for this database contains:

- A. Access to the most popular statistical packages;
- B. Disk access to selected databases;
- C. Utilities for moving offline files (most likely located on tape) online if needed for classes, professional/research use, etc.;
- D. Utilities for locating suitable data for analysis, depending upon the subject in which the investigator is interested;
- E. Consultation services, training, documentation preparation, membership fees for data services, purchase of miscellaneous data base files.

Each of these components is now described in detail:

A. Access to Statistical Packages.

The major statistical packages should be adequate for most uses. The minimum level of service should, in the view of the Social Science Research and Instructional Council, be access to SPSS and SAS. Many campuses have found SPSS historically to be extremely popular and adequate for most applications. Several data bases, including some developed by faculty, have been created and are available in SPSS system file format. Faculty sometimes also share their data in SPSS format.

A similar argument can be made with reference to SAS. It is less popular in the CSU system, but this phenomenon is largely because it was provided a decade later to CSU (and on a platform where it ran extremely inefficiently) than to the rest of the social science community in the nation. Many databases from the 1990 Census will be provided by the Census Bureau in SAS format, which is a further encouragement for users who process these files to become familiar with SAS. Many of the most advanced database users are SAS users.

There are other packages also becoming popular (the CSU Social Science Research and Instructional Council has had workshops on STATA, which some faculty use), and more specialized packages for specific uses not covered well in the major packages (for example, logit and probit regression). Many campuses will find that faculty will desire such programs.

B. Disk Access to Selected Databases.

Providing access usually assumes that the files reside on a storage medium to which faculty and students can gain access. Disk storage, due to its flexibility and processing speed, is usually the preferred choice. A class of 20 students cannot gain simultaneous access to a file if it resides on magnetic tape, which is usually the storage medium in which data bases are obtained from ICPSR and the other archives.

However, disk storage has often been impractical for infrequently used data bases because of its cost. Some files, for example, require over 1 gigabyte of storage. When the required amount of disk storage is not available, large or infrequently used files must remain stored offline on magnetic tape and a utility provided to move them from tape to disk. With the largest files, utilities are needed to reduce their size in the process of going from tape to disk.

Which files are provided on tape and which on disk will depend on faculty needs and the resources available on each campus. However, as a general guideline, the most frequently used files will require a minimum of 200 to 300 megabytes of disk storage.

C. Utilities for moving offline files online.

IRT has a number of utilities that faculty and IRT staff use to move files from tape to disk and vice versa. Since these have been written or provided for the Cyber, local utilities, depending upon the campus platform, will have to be used. If faculty are not familiar with the utilities or if disk space for faculty use is too limited, faculty will be dependent upon consultants' performing these activities.

D. Utilities for locating suitable data for analysis.

At present, IRT has a number of utilities available through which faculty and students can locate databases and questions of interest, usually by subject, but also by author, title, etc. The Guide to Resources and Services from the ICPSR (the 1,000 page book listing all databases and services available from the ICPSR) is available on the Central Cyber (ICPDB). CSUDB on the Central Cyber provides indexed searches by title or author; ICPNDEX does the same by subject. CSUVARS and QIKSTAT will provide a listing of interesting databases by variable content. QIKSTAT will not only assist in locating databases of interest but will do frequency distributions, scatterplots with bivariate regressions, and crosstabulations as well.

This area is one which is likely to be cut back for any campus which does not belong to a Social Science Speciality Center, if indeed one can be established. The costs of developing and maintaining these utilities are substantial, and they have all been written to operate on the Central Cyber. Their conversion, while a one time expense, is a substantial one.

Most faculty on most campuses will probably have to use the written Guide to Resources and Services, provided yearly from the ICPSR. Since there are no indexes or centralized listings for Field Polls and Roper surveys, access to these will probably be catch and catch can.

E. Consultation services, training, documentation, etc.

Support of the social science databases also includes costs that are not associated with the platform on which the databases are maintained. These include such things as:

- Consultation on data base access, information provided to faculty, students and staff on what data are available, could be ordered, how data are accessed, etc.
- Training workshops. CSU has had a number of training workshops over the years. These funds have been eliminated from the IRT budget, and given the difficulty incurred in placing them within the budget, it is VERY doubtful that they will be reestablished. Campuses wishing to have training workshops will probably incur little difficulty in having the standard short courses, which for the most part can be provided by local staff or even faculty provided with appropriate release time. But the longer one to two week workshops are gone for all practical purposes as part of the downsizing operation.
- Documentation preparation for both the statistical packages (now being done at most campuses) and the databases.
- Membership fees for data subscription services. Discussed below.
- Purchase of miscellaneous data base files. Discussed below.

V. Costs campuses will incur to provide these services.

These vary campus by campus, depending upon the patterns of use of the particular group of faculty and students. We have taken considerable care to incorporate statistics which best reflect actual per campus use of the social science resources in estimating the per campus costs.

Where appropriate, campus cost of providing a particular service has been taken as a percentage of the cost currently encumbered by IRT for providing that cost to all campuses. The actual percentage used in the computation is based on estimated campus usage of the resource. For example, IRT currently provides technical consultation regarding social science data base access at an annual cost of \$12,000. If a campus has utilized 10% of this resource, the estimated campus cost of providing the same service has been set at \$1,200.

Because each campus is characterized by different patterns of use, we have used two different measures of use in this report. IRT provided both; the data for each measure appear in Appendix **A**

System Usage: this is a measure of the actual computing requirements of each campus. It is based upon the actual SRUs of use of SPSS based upon the approximately 34% of SPSS usage which in a one year audit proved to be due to social science data base access. Thus campuses rating high in this category are those which show a high usage of SPSS on the Central Cyber in conjunction with social science data base processing on the Cyber. We will term this measure "system usage."

Research Usage: the second measure is based upon one year's worth of accesses to CSUDB, which, in the experience of the social science consultants at IRT, is a good estimate of the campus need for technical consultation and assistance. Because faculty research projects generate the most demand for

consultation, this measure is referred to as "research usage" and seems to be a good measure of research usage.

High "system usage" is generally associated with a propensity to use the Cyber 960 for classroom instruction; high "research usage" reflects the high need for individual consultation required for research and professional activities for CSU faculty.

A note on costs. Many of the costs associated with providing a social scientific data base library are not dependent on actual usage. Examples of such costs include membership fees for data subscription services, the purchase of data base files, and the development of utilities which enable students and faculty to gain access to the data. In these cases, the particular item will be associated with a fixed cost.

VI. Itemized Costs for Each Campus To Provide Access to the Social Science Data Bases

Costs have been divided into five categories as per the chart on the next page. A discussion of the meaning of the full and minimum models follows:

Full model:

Data Subscription Membership Fees: Under this model, the campus joins the relevant data base organizations, and it is assumed that the campus will provide a minimum level of Census data. The level of Census data is about the current average for CSU campuses (i.e., we are explicitly assuming that the total amount of Census data used by the larger campuses presently will not be provided -- see example at the end of this report).

Staff Support: I assume that, where appropriate, each campus will assume the prorated share of the staff support now provided by IRT. These amounts totally sum to approximately one professional position.

Data Storage: The full model assumes that hard disk space will be provided or purchased in sufficient amounts so that data sets of reasonable size (i.e., all but the largest Census files) can be left on line while classes and individual faculty are using them and kept off line on tape otherwise. Sufficient tapes are purchased to order datasets.

Software: already provided. Additional software beyond the requirements of SAS and SPSS may be desired by individual faculty.

Miscellaneous Hardware and Computational Requirements: full service provided in this area also.

Minimum model:

Data Subscription Membership Fees: Under this model, the campus joins the ICPSR and provides some Census data, but there is no membership in the Roper Center and fewer miscellaneous data base files are provided.

Staff Support: Instead of assuming its prorated share of the staff support now provided by IRT, this model assumes that these functions are basically absorbed

Social Science Annual Database Costs for Individual CSU Campuses

Full Minimum

Data Subscription Membership Fees:

Membership in the ICPSR	\$5,500	\$5,500
Membership in the Roper Center	\$2,000	----
Acquisition of Field Polls	\$0	\$0
Census of Population and Housing	\$2,200	\$2,200
Miscellaneous Database Files	\$2,000	\$1,000

Staff Support:

Computer Operator	\$1,500	----
Systems Analyst	\$1,500	\$1,500
ICC -- Technical Documentation	\$600	\$600
ICC -- Data Base Installation	\$15,000	\$7,500
ICC -- Subsetting/Downloading Files	\$450	\$450
ICC -- Develop/Maintain Utilities	\$12,000	----
ICC -- Prepare Documentation	\$6,000	----
ICC -- Training Workshops	\$150	\$150

Data Storage:

Magnetic Tapes	\$1,000	\$500
Disk Storage	\$3,000	\$3,000
Tape Drives	\$0	\$0

Software:

SPSS	\$0	\$0
SAS	\$0	\$0
Other	?	----

Miscellaneous Hardware:

Computational Requirements	\$6,500	\$2,000
----------------------------	---------	---------

Total Annual Cost: \$59,400 \$24,400

A detailed description of each itemized cost for the full service model is presented in the pages to follow.

by the present staff, with costs maintained only for a small amount of addition personnel costs. The assumption here is either that the current staff are sufficiently underworked that they can provide these services along with their current duties or that other duties will not be performed. Built into this model is a smaller level of data base acquisition.

Data Storage: The minimum model assumes that hard disk storage will be provided but that a small number of datasets will be ordered.

Software: already provided. No additional software will be acquired.

Miscellaneous Hardware and Computational Requirements: minimum service is assumed in this area.

Itemized Annual Costs For The Full Service Model

The pages that follow contain an itemized breakdown of the costs associated with providing full access to the social science data bases. Each item, and its cost, is described in detail. In cases where the cost is dependent on campus usage, a figure associated with average usage will be entered. With 20 campuses, average usage will be 1/20 of total usage or five percent (5%). A table displaying the projected costs for each of the 20 campuses given their actual usage of this resource appears in Appendix A.

Data Subscription Membership Fees

ICPSR - \$5,500

ICPSR is by far the largest supplier of data to the CSU. For an annual fee, ICPSR will supply virtually any of its holdings on magnetic tape. The CSU currently holds federated membership. This means that any campus can use the data but there can only be one person who orders and distributes the data. In order for a given campus to acquire data without IRT assistance, it will have to obtain a separate membership. The cost for campus membership is \$5,500.

Roper Center - \$2,000

Roper, like ICPSR, maintains an extensive data library. As a member, the CSU is kept abreast of new additions and can acquire a fixed quantity of data at no charge. The fee for membership to this service is \$2,000.

Field Institute - \$0

The CSU is currently a federated member of the Field Institute which administers public opinion surveys in California. Campuses will continue to receive Field survey data at no charge because the contract will continue to be centrally maintained. Field has generously offered to accommodate the decentralization process by distributing data to all campuses at no extra charge.

Database Purchases

Census of Population and Housing - \$2,200

None of the data subscription services supplies Census data without additional charge. Currently, the cheapest source is the State of California Department of Finance. The cost is \$165 a reel. IRT currently houses more than 100 reels of tape for the 1980 census. With the release of the 1990 tapes, students and faculty will need to gain access to these data. A collection of 40 tapes (20 for PUMS, 20 for STF) will likely serve most research and instructional needs. The cost is \$6,600 which prorated over a three year period comes to an annual cost of \$2,200.

Miscellaneous Data Base Files - \$ 2,000

IRT currently invests approximately \$2,000 per year acquiring data from miscellaneous sources including several state and federal agencies. Examples include INS data on immigration and BCS statistics on crime in California. These files are supplied at cost. Given that the files have been available to all campuses, and used by a great many, the cost is not prorated. They must be acquired if comparable support is to be maintained.

Staff Support

Maintenance of a data base library is generally handled by personnel falling into three distinct job classifications. These include an instructional computing consultant, a systems analyst, and a computer operator. The annual costs associated with employing full time personnel in each of these classifications were estimated at \$60,000 for an ICC, \$60,000 for a systems analyst, and \$40,000 for a computer operator. Given that only a portion of a staff position is actually required to support a given function, only the respective portion is used in estimating the cost of the service. Responsibilities of the ICC staff position are itemized separately.

Computer Operator -- \$ 1,500

Most of the social science data supplied to the CSU is supplied on magnetic tape. When these data are loaded for processing or for transfer to another storage medium, a computer operator must be present to mount the tape. A computer operator, or someone in a similar job classification, must also catalogue the tapes as they are acquired. IRT currently dedicates approximately 3/4 of a staff position in this classification to support access to the social science databases at a cost of \$30,000. Campuses can expect to bear comparable costs yet only in direct proportion to their usage (combined measure) of this resource. For the average campus, this cost is \$1,500. Although not reflected in the table to follow, each campus should anticipate a minimum cost of \$1,000 in this classification to support data installation. ($\$30,000 \times .05 = \$1,500$)

Systems Analyst -- \$ 1,500

IRT currently dedicates approximately 1/2 of a staff position in this classification to support access to the social science data bases. Responsibilities include maintaining the various operating systems and providing consultation regarding their use. This service is provided at an approximate cost of \$30,000. The estimated campus cost for providing the same service is taken as a proportion of the total cost with each portion being prorated on the basis of actual usage of this resource (combined measure). ($\$30,000 \times .05 = \$1,500$)

ICC Technical Consultation -- \$600

Two types of consultation are included under this heading - consultation regarding the identification of relevant data and consultation regarding actual processing. IRT currently allocates 1/5 of a staff position to perform these functions at a cost of \$12,000. Campus cost estimates are based on the share of usage (research measure) associated with this resource. A typical 5% usage pattern is associated with a \$600 cost. ($\$12,000 \times .05 = \600)

ICC Data Base Installations -- \$15,000

To maintain a social science data base library as comprehensive as that currently maintained by IRT, support staff must anticipate faculty and student needs and acquire data appropriately. The data must then be installed. IRT currently allocates 1/4 of a staff position to perform this function at a cost of \$15,000. Actual responsibilities include ordering, loading, and testing data, preparing data base abstracts, formatting data when appropriate, archiving codebooks, and cataloging all relevant access information. The costs associated with this function are not prorated given that each campus will be required to perform this function if current levels of access are to be maintained.

ICC Subsets, Downloads Files -- \$450

IRT currently allocates 15/100 of a staff position to extract subsets of data base files used in faculty research and for classroom instruction. These files are sometimes downloaded for PC or MAC use. The estimated campus cost of providing this resource is prorated using the campus share of usage of this resource (research usage). ($\$9,000 \times .05 = \450)

ICC Develops, Maintains Access Utilities -- \$12,000

Responsibilities contained under this heading include the development and maintenance of utilities for printing online codebooks (i.e. data base documentation files), procedures which enable searches of data base holdings, and programs which simplify accessing data base files. Reference IRT's CBOOK, CSUDB, CSUVARS, ICPNDEX, QIKSTAT, DBTAPE, and TAP2DSK. The costs associated with this function are not prorated given that each campus will be required to perform this function if current levels of support are to be maintained. IRT currently allocates approximately 1/5 of a staff position to provide this resource at a cost of \$12,000.

ICC Prepares Documentation -- \$6,000

IRT allocates approximately 1/10 of a staff position to prepare documentation regarding social science data base access at a cost of \$6,000. Given that documentation is generally installation dependent, each campus will need to prepare its own documentation regarding data base access and the utilities which are provided. The cost is not prorated.

ICC Assisted Training Workshops -- \$150

IRT has routinely provided staff to assist in systemwide faculty training workshops related to data base access. Approximately 1/20 of a staff position has been allocated at a cost of \$3,000. The cost that each campus will bear in providing comparable training is prorated at exactly 5% usage per campus. Systemwide workshops generally include one faculty member from each campus. ($\$3,000 \times .05 = \150)

Data Storage**Magnetic Tapes -- \$1,000**

Approximately 50 additional tapes (2400') are required to meet the storage requirements for newly acquired data in a given year. (50 reels @ \$20 = \$1,000)

Disk Storage -- \$3,000

The most popular social science data base files require about 200 megabytes of disk storage. An average campus will also need approximately 200 megabytes of storage for customized subsets of larger files and yet another 200 megabytes to be used as a work area for installing and subsetting data. Perhaps as much as an additional 600 megabytes will be needed for the personal accounts of students and faculty. The estimated disk storage which each campus should anticipate providing is somewhere between .6 to 1.2 gigabytes. The estimated cost of providing only .6 gigabytes is entered above. This cost is taken as \$ 15,000 and prorated over a five year period for an annual cost of \$3,000. The \$15,000 amount is the approximate cost of installing a DEC RAS2 622MB drive. Campuses showing high usage (system usage measure) should consider adding a full gigabyte of storage. However, only the \$3,000 amount is used in computing the cost to each campus.

Tape Drives -- \$0

Development, maintenance, and support of a social science data base library will probably not place severe strain on the magnetic tape drives unless there are some remarkable changes in usage patterns. Nonetheless, they are a requirement, and each campus should maintain a minimum of two such drives for this purpose. Given that most campuses already support such drives, a \$0 amount is shown entered above.

Software**SPSS -- \$0**

Given that campuses will likely maintain licenses for SPSS to support a variety of functions, no additional cost will be entered in conjunction with its use to support social science data base access.

SAS -- \$0

See comment regarding SPSS above.

Miscellaneous Hardware**Computational Requirements -- \$6,500**

The costs summarized above account for approximately \$130,000 of IRT's total estimated annual expenditure of \$260,000 to support access to its social science data base library. Nearly all of the remaining \$130,000 is used to maintain the hardware used for actual computation (i.e. a share of the Cyber 960). In a decentralized environment, each campus will need to upgrade their own hardware to accommodate faculty and students who formerly used the Cyber 960 to process social science data bases. Estimating the cost of providing this service is difficult given the diversity of computing environments. However, for the purpose of this

report, the cost will be estimated using a prorated share of the \$130,000 figure. The campus share is based on actual system usage not research usage of this resource. A typical campus share of 5% usage would be replaced at a cost of \$6,500 ($\$130,000 \times .05 = \$6,500$).

Total Cost Per Campus* -- \$58,400

* The cost shown above is the estimated annual cost a typical campus will bear in providing these services. Estimated costs for each campus, given their distinctive usage patterns, are presented in Appendix A.

VII. Conclusions.

1. Costs.

The costs of providing relatively full social science data base access for each campus in the CSU average \$58,900 per campus. For twenty campuses, this totals to \$1,178,000. Since the funds being returned to each campus for the social sciences alone total only a quarter of that amount, it is clearly not feasible for every campus to provide social science data base services at this level.

2. Service.

A minimum level of service -- much less and probably much slower than at present -- for each campus can be provided for about \$25,000 per campus. This amount is twice the \$13,000 being provided the average campus for the social science functions.

3. Service Declines.

Since the functions cost more to provide than the average campus will receive under either scenario, the most likely conclusion for 1991-1992 is that services to faculty on the average campus will decline. In some cases the decline will be substantial; for perhaps a few campuses, it will be less so.

4. A Social Science Speciality Center?

It would make more sense for all or most campuses to join a social science speciality center. But under the current "rules" adopted by IRT's "Committee of the Whole," voluntary taxation is the funding scenario to be used. In working out the implications of voluntary taxation, we have hit on some substantial problems: planning for the future when a campus can decide not to pay at the last minute, what to do about faculty who wish information and come from a campus which has not joined, etc.

These are substantial enough that as of this point (just before New Year's Day), it does not appear that there will be a Social Science Speciality Center proposal from any campus.

The major recommendation of this report is that CSU consider providing "off-the-top" funding to a social science speciality center in order to provide for an appropriate transition and to preserve the current level of faculty services

provided by IRT. My conclusion is that the level of faculty services will decline substantially -- indeed, it has already declined -- unless this action is taken.

5. The Centralized ICPSR Membership.

Under the scenario where there is no centralized social science speciality center, each campus will be left to join ICPSR or not. The centralized funding that has been provided for ICPSR membership (in the Library budget -- not the IRT budget that is being reorganized) at half the cost of each campus' joining for \$5,500 will become superfluous and can -- indeed, should⁵ -- be eliminated from the budget.

6. A Note on Why Decentralization Does Not Seem To Work Easily With the Social Science Data Bases OR Why It is Likely That Service Will Decline.

The social science data base library is a shared and integrated resource. It is not comprised of units which can be easily separated and dispersed. Providing access to the census data is an excellent case in point. In a decentralized environment, the campuses together could spend a total of over \$100,000 to acquire small and relatively incomplete collections of these data. A systemwide resource, on the other hand, can provide access to virtually all of the needed files at a cost of only \$20,000. Similar cost savings are also realized in the systemwide support and development of utilities that facilitate access to these data.

7. What Will \$13,000 Provide?

An inspection of the itemized costs presented in this report provides an answer. It can purchase campus membership to the ICPSR (\$5,500), the computing power equivalent to one IBM 286 or 386sx clone (\$2,000), an inadequate collection of census data (\$500), and 1/12 of a staff position (5 hrs/week @ \$5,000/yr) to order, install, catalogue, and facilitate access to the greatly diminished number of data base files which can be accommodated.

8. What Good Is Access to Social Science Data Bases?

The practice of research in the social sciences is not an idle exercise aimed at generating purely theoretical abstractions. It is practiced in order to gain a better understanding of our world and to make informed decisions about our future. Gaining timely access to current or historical information is a critical component.

Here are two examples of how access to relevant data provided through a data archive has filled an important role both for the system and for the State of California. On one occasion, a faculty member, with the help of her students, used such information to generate a report describing the geographic distribution of senior citizens who had no means of transportation. This report was used as a needs assessment for a program to provide for the elderly. Another faculty member, again with student assistance, has used court sentencing data to evaluate the effectiveness of a particular rehabilitation program. Obviously, students benefit by having the opportunity to engage in research at this level.

⁵ There is no point in having a centralized membership for the ICPSR without a centralized point of contact for the ICPSR membership. If the campuses are to decide whether to join individually, centralized membership is irrelevant.

Appendix A

The table below summarizes the estimated annual costs each campus will bear in providing comparable levels of access to the social science data bases in a decentralized environment. Estimates are based on usage characteristics of each campus. The table also summarizes the actual funding being returned to each campus in conjunction with the decentralization of this function.

<i>Campus</i>	Estimated Share of Research Usage in Percent	Estimated Share of System Usage in Percent	Combined or Average Measure of Usage in Percent	Percentage of Funding Returned to Campus	Respective Funding Returned to Campus in Dollars	Cost to Campus to Replace Services in Dollars	Total Annual Savings per Campus in Dollars
Bakersfield	1.3	.7	1.0	1.3	3,335	50,633	- 47,298
Chico	.9	1.0	1.0	5.0	12,875	50,909	- 38,034
Dominguez Hills	.3	.2	.3	2.2	5,738	49,323	- 43,585
Fresno	2.6	14.5	8.6	5.8	15,080	73,376	- 58,296
Fullerton	7.2	13.7	10.5	6.8	17,759	74,442	- 56,683
Hayward	3.0	18.9	11.0	3.6	9,249	80,620	- 71,371
Humboldt	1.4	1.2	1.3	2.1	5,578	51,484	- 45,906
Long Beach	7.4	6.7	7.1	8.2	21,248	63,344	- 42,096
Los Angeles	2.9	3.7	3.3	5.2	13,612	56,249	- 42,637
Northridge	20.9	6.1	13.5	7.6	19,769	69,269	- 49,500
Pomona	8.2	3.7	6.0	7.5	19,384	58,952	- 39,568
Sacramento	12.2	.8	6.5	7.4	19,211	56,352	- 37,141
San Bernadino	2.1	9.8	6.0	2.9	7,517	65,601	- 58,084
San Diego	6.4	5.2	5.8	9.5	24,733	60,434	- 35,701
San Francisco	14.6	3.3	9.0	7.2	18,606	61,576	- 42,970
San Jose	2.1	1.4	1.8	8.0	20,695	52,161	- 31,466
San Luis Obispo	2.0	6.3	4.2	6.6	17,148	59,950	- 42,802
San Marcos	.0	.1	.1	.1	247	49,010	- 48,763
Sonoma	3.4	.0	1.7	1.7	4,507	50,584	- 46,077
Stanislaus	1.2	2.5	1.9	1.4	3,711	53,462	- 49,751
Total	100%	100%	100%	100%	260,000	1,188,000	-928,000