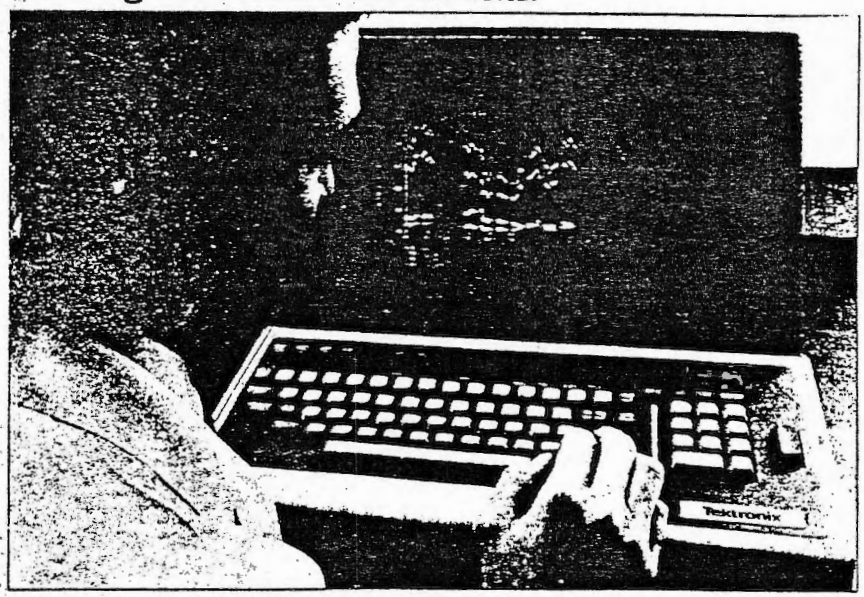
Examples of RS/1 Output

Seeing scientists' schemata



Research scientists will be able to graph and interpret their data promptly with the use of RS/1, a software research system developed by Bolt Beranek & Newman, Inc. Used in conjunction with a minicomputer, RS/1 presents a visual representation of pertinent data on a graphics terminal. The decision-making process is accelerated because scientists no longer have to wait overnight for processing of data. RS/1 combines the functions of data management and statistical analyses, and has full graphics capability. The types of statistical analyses available include analysis of variance, linear and nonlinear regression, and parametric and nonparametric testing. It is available for the full range of DEC PDP-11 and VAX computers and can support a variety of Tektronix and DEC terminals, including color graphics. Bolt Beranek & Newman are marketing RS/1 themselves. It costs $15,400. BBN is located at 50 Moulton St., Cambridge, MA 02238, (617) 491-1850.

ARMA 8C X 21R ARMADILLO TAIL-GRO DATA

	5 INITIAL TAIL(CM)	6 DELTA TAIL(CM)	7 GROWTH/ DOSE RATIO	8 PERCENT INCREASE
1 A352C	5 5	2 4	0 292683	43 6364
2 A681D	20 9	3 4	0 257576	16 2679
3 A785R	21	5 3	0 289486	25 2381
4 A852R	6	2 2	0 278481	36 6667
5 A965D	27 4	8 1	0 224377	29 562
6 A572C	4 9	2 8	0 291667	57 1428
7 A631D	23 3	4 2	0 24	18 0258
8 A536R	25 2	5 8	0 234818	23 0159
9 A867R	26 9	6 9	0 221154	25 6586
10 A945C	6 9	2 9	0 325843	42 029
11 A856R	5 1	2 7	0 290323	52 9412
12 A978D	27	5 9	0 282749	21 8519
13 A877R	31	3 1	0 289459	10
14 A938C	28 1	7 5	0 287182	26 6984

Input to and output from Bolt Beranek and Newman's data management system is based on the two-dimensional data table format common in research

System Capabilities

Command Language

RS/1 operates through a simple command language that scientists can readily learn. RS/1 commands such as

MAKE TABLE
or
TEST NORMAL COLUMN 1 OF table

make highly complex operations accessible through English statements.

Supplemented by a built-in programming language, the RS/1 command language gives the user access to a full range of data handling techniques, including

- Data Entry and Retrieval
- Data Transformation and Analysis
- Graphics
- Statistical Analysis
- Curve Fitting
- Applications Programs

Data Entry and Retrieval

Data is stored and used in the format of two-dimensional tables. The user can create hundreds of tables, and use any number of them at one time. The rows and columns—up to several thousand per table—can hold large amounts of number, text, date, or "EMPTY" data. Operating on full tables, on selected rows or columns of various tables, and on portions of tables screened by data value—e.g., WHERE COL 8 = "CONTROL"— the user has a wide range of data handling tools:

- MAKE TABLE from manual input, ASCII files, binary files, other tables, computational results
- DISPLAY table or table portion
- CHANGE table data, ADD COLUMNS, ADD ROWS, INSERT columns or rows
- SORT, MERGE, COMPARE, MAKE TABLE AS DISTRIBUTION, TRANSPOSE table

Data Transformation and Analysis

Data in tables can be used to derive values through mathematical expressions. An entire column, for instance, can be derived from one or more other columns with a single command, such as SET COL 2 OF TBL TO LOG (COL 1). Conditional expressions may be incorporated into computations to select rows or columns according to values they contain—e.g., RESPONSE = MEAN OF COL 8 WHERE COL 2 NE 0. Such transformation and analysis expressions can use table data, numbers and variables with a full range of operators, and all functions available in the programming language (see reverse):

- Arithmetic +, −, *, /, ** (exponent), unary + and −
- Logical AND, OR, NOT
- Grouping ( )
- Relational EQ, NE, LT, LE, GT, GE
- Text CAT and Date +, −

Graphics

Graphs, like tables, are created through a simple interactive dialog or though a single command, and are saved as data objects. Commands and procedures—such as MAKE GRAPH, ADD CURVES, BARGRAPH—provide extensive capabilities to make, modify, add to, and display:

- Scatter diagrams
- Line graphs
- Histograms
- Bar graphs
- Color graphics

Command-level graphics capabilities:

- MAKE GRAPH from table portions, manually input values, directly plotting any functional form
- ADD and DELETE data or function curve, mean/standard deviation curves
- Modify data format, SET X or Y AXIS bounds, data SYMBOL, SET LABEL of axes and legend
- Call to procedures—COPY, ERASE, draw-line

Graphs are also used extensively in statistical analysis.

Statistical Analysis

An extensive set of statistical tests can be applied to data in tables. Simple commands—such as TEST MEANS of table portions—invoke RS/1 statistical tests and analyses:

- Display of any or all descriptive statistics:
  SUM, MEAN, VARIANCE, STDEV, DEGREES, MOMENT2, MOMENT3, MOMENT4, SKEWNESS, KURTOSIS, MEDIAN, MAXIMUM, MINIMUM
- Wilk's test for normality of small samples
- Kolmogrov-Smirnov test for normality
- Numerous t-tests (one- and two-tailed)
  mean of one sample
  means of two samples, paired, pooled, or unpooled variance
- Variance ratio F-test
- Nonparametric tests
  Wilcoxon test (rank sum technique)
  Sign rank test
  Rank correlation coefficient (Spearman's rho)
  Ansari-Bradley test of dispersions
- Build contingency table
- Chi-square test (frequency data)
- Analysis of variance, one- or two-way
- 2K factorial design analysis of variance
- Analysis of covariance (one-way)
- Correlation coefficient

Curve Fitting

Regression analysis is performed on data in tables or graphs. A single command, such as FIT LINE TO CURVE of graphs, returns the fitted result and numerous goodness-of-fit parameters.

- Linear, polynomial fit
- Full nonlinear regression analysis of any functional form, e.g., multi-term exponential
- Marquardt-Levenberg method for nonlinear functions
  no need to input derivatives
  optional weights
  control over convergence

Regression and variance analysis automatically return output tables showing, as appropriate:

- Coefficients, standard deviations, t-values and significance levels
- Analysis of variance including F-value and significance level
- Residual values
- Standard deviation of regression
- Multiple correlation coefficient

Applications Programs

Additional analyticial capabilities are available in programs provided with RS/1. These built-in procedures include applications-oriented analytical packages:

- Solution of simultaneous linear equations
- Dissociation constant for agonist drug, antagonist drug
- Least-squares fit to logistic function
- Symbolic differentiation

A bridge to the BMDP statistics package gives BMDP licensees access through RS/1 to extensive and specialized statistical operations.

Programming Language

RPL—The Research Programming Language—enables RS/1 users to extend and tailor system capabilities for their applications. RPL is fully integrated into the RS/1 data management system, and can be used to develop procedures through:

- Direct use of all RS/1 commands and procedures
- Direct access to RS/1 tables and graphs
- Additional table-handling functions and support for very large arrays
- Structured programming statements
  IF-THEN-ELSE,
  DO WHILE and indexed loops,
  PROCEDURE-CALL-RETURN, optional arguments
- Interactive tools for user input
- Error-handling mechanisms
- Procedure editor, compiler
- CHAIN to programs outside RS/1 (e.g., FORTRAN programs)
- Access to ASCII or binary data produced by other packages and languages
- ENCODE and DECODE formatted data
- Noninteractive access to statistical tools (stored results), e.g., CALL PUBLIC CORRELATE

RPL Functions

RPL functions can be used in RS/1 commands as well as in user procedures.

- Trigonometric
  SIN, COS, TAN, ASIN, ACOS, ATAN, SINH, COSH, TANH, SIND, COSD, TAND
- Datatype
  Test for number, TYPE, EMPTY, NOTEMPTY
- Numeric
  INTEGER, FRACTION, CFIXED, MOD, RNDOFF, SIGFIG, ABS, FLOOR, CEIL, SHIFT, SQRT, EXP, LOG, LOG10, MAX, MIN, LOGAND, LOGOR, LOGNOT
- Character
  CHARS, CAT, LOC (search), EXT (substring), CAPS

Appendix E - IMSL

IMSL consists of callable Fortran subroutines to perform certain mathematical and statistical functions. A listing.of the various subroutines follows.

INTERNATIONAL
MATHEMATICAL &
STATISTICAL
LIBRARIES, INC.

March 11, 1981

Mr. James Billingsley
Dept. of Commerce - OFSPS-DIDS
2001 S. Street NW
Rm. 200
Washington, DC 20009

Dear Mr. Billingsley:

It was a pleasure to speak with you about the IMSL Libray.

Any organization involved with the development of scientific application programs should be a subscriber to the IMSL Library. This high quality software is comprehensive, well documented, maintained, supported, and inexpensive. The subroutines are used as "building blocks" in program development to save costly man-hours. Furthermore, the time tested algorithms developed by leading authorities in their field prevent errors which can be so destructive.

I have enclosed the IMSL Library Information brochure for your review. It presents a general overview of Edition 8 of the IMSL Library and describes its contents. It also gives examples of the documentation supplied with each ability.

Our subscribers (currently over 1200) have found that the annual subscription to our Library has been highly cost effective in the areas of improved computational accuracy and more efficient utilization of personnel in scientific, engineering and research application development.

I have also enclosed IMSL order form F-S008 which describes the pricing, terms and conditions associated with a subscription to the IMSL Library. The VAX 11/780 and PDP 11 computers are classed as Type II computational systems. Therefore, the annual subscription rate for your computer is $1,400. Discounts are offered to degree granting educational institutions and for multiple year subscriptions.

If you would like to be added to our mailing list, fill in and return the enclosed postcard questionnaire. Please let me know if you require further information.

Yours truly,

Janet Erbacher
Marketing

JE/mk
Enclosures

NBC Building, 7500 Bellaire Boulevard, Houston, Texas 77036-USA
Telephone: (713) 772-1927 Telex 79-1923 IMSL INC HOU

STATISTICAL FEATURES

DATA DESCRIPTION AND SCREENING

Means, standard deviations, coefficient of variation, skewness, kurtosis, median, mode, four robust estimates of location, range, interquartile range

Breakdowns—univariate statistics for each cell defined by one or more cross-classification variables

Data printouts—all data, data after transformation, ordered data, standard scores, only cases with extreme or missing values, residuals from regression, scores from factor analysis, principal components or canonical variables

Histograms, cumulative histograms, normal probability plots

Bivariate plots

Missing value pattern display

FREQUENCY COUNTS AND TABLE ANALYSIS

One-way frequencies of all distinct values or frequencies of values rounded by specified amounts

Two-way frequency tables, frequency tables for levels of one or more additional variables; tallies of missing or extreme values on two-way table marginals

Multiway frequency tables with optional display of all possible marginal tables

More than twenty measures of association for two-way tables with appropriate standard errors for hypothesis testing and confidence intervals

Stepwise identification of extreme cells in two-way tables

Log-linear analysis of multiway tables with aids for model building and orthogonal decomposition

ANALYSIS OF VARIANCE

One-way and two-way analysis of variance with side-by-side histograms—unequal cell sizes; both Brown-Forsythe and Welch tests to allow unequal cell variances

One-way analysis of covariance with many types of residual plots and tests for equality of regression slopes

Multifactor analysis of variance and covariance—including factorial, repeated measures, fractional factorial, Latin square, unequal sample sizes, multiple covariates, etc.

General mixed model unbalanced analysis of variance and covariance by maximum likelihood and restricted maximum likelihood

General mixed model analysis of variance for equal cell sizes—highly efficient analysis (New)

REGRESSION

Comprehensive coverage

Simple linear regression with regression line marked on frame of plot

Multiple regression for whole sample and subgroups

Stepwise multiple regression—forward and backward stepping

All possible subset regression using efficient Furnival-Wilson algorithm

Regression on principal components

Multivariate regression

Nonlinear regression and maximum likelihood estimation

Stepwise logistic regression (New)

Special regression features

Extensive plotting capability is an integral part of regression—predicted values, observed values and residuals vs. independent variables; normal plots; residuals and squared residuals vs. predicted values

Standard errors for predicted values

Properly standardized (Studentized) residuals

Multivariate outlier identification

Cross validation in regression

Derivative-free nonlinear regression

MULTIVARIATE TECHNIQUES

Cluster analysis of variables; hierarchial cluster analysis of cases (four distance measures); K—means cluster analysis; simultaneous cluster analysis of cases and variables

Principal components

Factor analysis—maximum likelihood, several types of factor rotation, many graphical displays

Stepwise discriminant analysis—Lachenbruch and Mickey (jackknife) method of evaluating correct classification; classifies new cases not used in the computation; canonical variable plots

Partial and canonical correlation

OTHER FEATURES

Life table analysis—estimates survival distribution of individuals observed over varying time periods

Nonparametric statistics—for data that deviate too much from the usual analysis of variance assumptions

TABE

March 28, 1981

MEMORANDUM FOR: Joseph W. Duncan
Director, OFSPS

FROM: James B. Billingsley
Technical Manager, DIDS

SUBJECT: Recommended Systems Enhancements

The term "operational" has been applied to DIDS as early as last September when the systems were installed in their current locations. In the DIDS Management Plan for 1981 (page 5, para. A), the need for the "operational" data base is addressed. In the following paragraphs, the problem areas that must be resolved before we can support DIDS as, what I would consider, an "operational" system are presented. The four main problem areas discussed are: (1) the inaccurate and incomplete map data base; (2) the new "operational" variable data base; (3) the lack of analysis and business graphic capabilities; and (4) the lack of back-up facilities.

Less important enhancements are not detailed since I feel that reaching operational status will require more than our current resources can support.

DIDS Maps

Problems: The DIDS maps were prepared using software written in 1978 for the original White House demonstration. Since the early effort was to demonstrate terminal feasibility, accuracy was not very important. As a result, many census tracts are missing from the DIDS maps and many more incorrectly numbered. This problem may also be present in the state maps and possibly even in the U.S. map as well. To date, approximately 20 (1970) SMSA maps have been installed on the system. An additional 260 SMSA maps are available on tape ready for conversion. All will suffer the same fate upon installation using current software. Digitizing of 1980 maps is incomplete and potentially may not be performed unless funds are provided to the Census map generation group.

Solution: Since the DIDS base maps represent the common translator for all displayed variables, it is essential that each map be validated for accuracy after installation and its quality certified. Validation and certification should be performed interactively by the display of maps on the DIDS monitor and the comparison be made tact by tract and county by county between the displayed map and the equivalent paper map standard. This problem is solved by the development of new map input software, assignment of people to perform the validation, and providing funds for 1980 map generation. Manpower and cost estimates are as follows:

1. Map generation and validation software $7K;
2. Map validation, 1 to 1/2 man years; and
3. 1980 map generation, five man years.

NOTE: The validation process will use DIDS terminal resources. Since considerable time is involved, this process will be performed on second and third shifts to minimize impact on normal DIDS operations. Support funding for additional shifts must be added.

DIDS Data Base

Problems: The current data base has been installed over the past 2 1/2 years as test case data sets predominately to stimulate interest in the DIDS program. None of the data sets have been validated after entry in DIDS nor has there been any effort to protect the data from inadvertent or deliberate modification.

The DIDS Data Base Working Group (DBWG) will shortly define the elements to be installed in the new operational data base. The operational data base must include documentation and keywords to support the data sets. We are in the process of assembling standards for the combining of data, documentation and keywords into a viable exchange format.

It is assumed that under White House pressure, the data base size will be increased significantly and that the demand will be for the new data base to be on-line and fully operational in a very short time. Any increase in on-line data base size will impact the hardware on both the remote and host systems.

Therefore, the problems associated with the data base can be listed as follows:

1. Lack of standards for data exchange;
2. Lack of system software support for documented and keyword search;
3. Lack of data validation processes;
4. Lack of data protection;
5. Insufficient capability to handle additional menu list at remote;
6. Lack of community support for Data Base Working Group; and
7. Insufficient storage at Host VAX 11/780.

Solutions: The DBWG is a key to the success of establishing an operational data base. Members must be assigned soon from the agencies or we must take alternative action to obtain data in the new format.

It would be desirable for the DBWG also to complete the standards document for the data exchange, however, again the lack of a representatives to the DBWG and limited internal staff preclude this activity. I, therefore, recommend that we contract The Mitre Corporation to assist with the following: (1) development of new operational data base standards documentation which includes data documentation and keyword standards; (2) develop a new data exchange format which minimizes the interchange problems associated with the current format due to differences between computer systems; (3) provide assistance to the DBWG through efforts to familiarize agencies with the new exchange format and to assist, where possible, with the actual exchange; and (4) study the potential for the addition of thesaurus and Boolean search capabilities on DIDS.

After standards and exchange formats are developed and cooperative efforts between agencies start data flowing into our facility, we must be prepared to enter, validate, protect, support keyword search, and store and display the

documentation, as well as the data. New software must be developed to support all of these functions including data base management software to link the menu, keyword, documentation, and the data sets together and provide rapid access from compressed storage. Therefore, a significant software development effort is required to accomodate the operational data base. If we assume that this work must progress at a rapid rate to support pending White House information access projects, the software development staff must be augmented to accommodate the additional work load. Recommended additional staff and costs are shown below.

When the operational data base (map and variable) is installed, we will rapidly fill the three disk systems on the host to capacity. The addition of documentation further complicates the storage problem due to the additional information needed to support each data element. Original thinking was to support the program with on-line and off-line data by replacing disk packs with data needed by a user. This approach is probably acceptable to most users, however, I believe the White House will want most data sets on-line for rapid access. The host system especially fits this need since the remotes call for data from the host in response to user requests. The only solution is to ad more on-line disk storage to the host computer system to accomodate the number of data elements likely to be required on-line. A weak first guess would recommend the addition of two additional disk systems. This would support up to 24,000 variables (with documentation) on-line on three 176 mbyte disks, a fourth 176 mybte disk for map storage and retrieval by users at the host terminal, and a fifth disk for system and applications software.

The disk storage at the remote site was also chosen to support reasonable numbers of map, variable, menu and software storage. The assumption being that for the low cost of the system, the user was not to expect all data, maps and functions to be available on-line.

I do not believe the White House will accept the delays and inconvenience created by significant disk swapping to analyze their data.

The remote communcations link supports variable data access (additional 10 seconds added to response) however, the current software and menu disks are full so that no additional menu elements or software features can be added without changing disks during operations. Moreover, the addition of documentation, and keyword search will further burden the storage requirements of the system. Again, the solution is the addition of more disk storage at the remote terminal to allow more maps, menus, documentation, variable data, and software functions on line. I, recommend that we add three 10 mbyte disks to the remote terminal to support the increased access needs and a 10 mbyte disk on the host computer to write the disks.

The following costs and manpower represent a initial rough approximation and must be revised as further details are worked out.

1.	Mitre support	$ 50K
2.	Additions to existing software support	$ 70K
3.	Two additional 176 mbyte disks	$ 90K
4.	Four 10 mbyte disks	$ 30K
	Improved data base total	$240K

Analysis and Graphic Generation

Problem: The DIDS computed variable function represents the limit of current DIDS analysis capabilities. Many users, and the steering committee have expressed the need for more sophisticated analysis functions including regression, variance, smoothing, sampling and some limited modeling capabilities. Display and hard copy production of analyzed results in bar chart, pie chart and three dimensional form is also required.

It would appear that the White House effort will raise these items to a much higher priority than originally planned creating significant programmatic problems due to our limited internal and contractor staff.

Solution: The most desirable approach to solving both the analysis and graphic problems would be to carefully define user requirements and then translate into software and hardware elements supporting the developed specification. I do not believe time will allow the luxury of a requirement definition study since the demand will press for immediate solution. I, therefore, recommend a short review of available analysis and graphic software systems commercially available which fit our most pressing problems (and our computer systems) and the purchase and interfacing of the best candidate software with our system. I also recommend the expansion of the VAX 11/780 memory system from the current 512K bytes to 2 mybtes to accommodate the additional working space needed for analysis and graphic functions.

A review of output product generation equipment is also necessary to produce paper products (graphic, map, and printout products). Either direct color Xerox or ink jet plotters have sufficient resolution to produce quality output products inexpensively. The computer printer is adequate for printed products, however, printer capability must be added at the remote terminal site if printed output is desired at the Old Executive Office Building.

Graphic and alphanumeric presentations (color and black and white) can be output on the Dunn camera in both 8" x 10" Polaroid and 35 mm slide photographic formats by the installation of a simple switch system on the Dunn camera. If these products are needed at the Old Executive Office Building, an additional camera system is required for that location.

Approximate costs for the analysis and graphic capabilities is presented in ranges of cost since a review of available products is necessary.

1.	Analysis software	$10K to $50k
2.	Graphic software	$15K to $35K
3.	Integration into our software	$45K to 55K
4.	VAX memory	$35K
5.	Output device	$30K to $80K
6.	Additional Dunn or Matrix camera	$20K
	TOTALS	$165K to $275K

Back-Up Facilities

Problem: Little or no back-up equipment is available to support component failure. In the original demonstration on Capitol Hill and the White House, I had a full duplicate of each piece of equipment within easy reach so that little time would be lost due to failure. I also had a large staff of technical support people stationed at each point of possible failure to quickly respond. Twice during the 1978 demonstrations, we had failures that were corrected almost instantly. The modem link for the alphanumeric terminal failed in the middle of a demonstration, however, with one phone call, service was restored in three minutes. The second failure was the Advent projection T.V. which was instantly replaced with the back-up unit by the standby technicians.

As you can see, we were prepared to support the demonstrations with maximum back-up to reduce or eliminate any embarrassment. I don't think that it's practical to provide the level of back-up provided for the 1978 demonstrations for day-to-day operations, however, we currently have such limited back-up facilities that to state we could support regular White House briefings is ludicrous.

I make no recommendations on needed back-up since I do not know how reliant the White House will become on DIDS availability. This problem must be explored in detail and funded appropriately.

Conclusions: The major elements discussed in this memo represent the basics of a DIDS operational system. Over $500K potentially is needed to properly support functions as outlined for DIDS operations support. These costs do not reflect the increased costs for back-up functions (equipment and people) nor do they reflect the costs of the system I proposed for the White House (Old Executive Office Building).

Their is an obvious need to discuss details concerning the level and quality of DIDS activity required for the future.

TAB F

May 14, 1981

MEMORANDUM FOR: DIDS Executive Committee

From: J. Timothy Sprehe
Chairman, Data Base Working Group

Subject: Recommendations arising from the Data Base Working Group

I. Initial List of Candidates for Operational Data Base

It is a pleasure to report that the pace of activities in the Data Base Working Group (DBWG) has quickened measurably. DBWG has met once in plenary since its reorganization and its three subgroups have met at least twice apiece. Agency representatives are interested, active and contributing significantly. Most recently (May 13), the Data Suppliers Subgroup and Data Users Subgroup met jointly to discuss specific candidates for the DIDS operational data base.

Attached is a list of the initial candidates put forward by group members. *This list is for information purposes only and does not constitute a set of recommendations by DBWG.* The reason DBWG does not yet recommend this list is that we are simply not quite that far in our work. It is quite possible that DBWG may make such a recommendation at the next meeting of Users and Suppliers on June 4, but the two subgroups believe they have not adequately reviewed and discussed the list as of the present moment.

The purpose, therefore, in presenting this list to you is (1) to satisfy your curiosity, and (2) to indicate the direction of our collective thinking.

Basically, DBWG has taken a very simple approach to choosing candidates for the operational data base. In the absence of any instruction or information to the contrary, we have felt the best predictor of what the National Indicators System and other users would want is what we colloquially call "agency best-sellers." That is, we have asked the agencies: for the first operational data base, would you please specify those data which your users most often request and which you believe to be your most valuable products. Since this is the first operational data base and we assume there will be additions, we have asked the agencies to specify an exclusive rather than inclusive data set.

II. The Need for Communication to the Agencies

Quite understandably, the members of DBWG ask: what happens to DBWG recommendations concerning the operational data base?

My answer has been the following. Assuming that the DIDS Executive and Steering Committees accept the recommendations of DBWG substantially as presented, then letters will go out from the White House to agencies which will request the agencies to prepare and submit to DIDS the list of data agreed upon. I have assumed that agencies will be required to submit these data (1) in DIDS exchange format, (2) accompanied by DIDS on-line data documentation requirements, (3) with appropriate keywords indicated, and (4) to accomplish the submission by August 1981. I have further assumed that agencies will be asked, in the same letter from the White House, to adopt procedures for routinely maintaining and updating these data on the DIDS system. Finally, I have assumed that agencies will be expected to comply with this request out of existing financial and personnel resources.

The members of DBWG can see that the White House request, when it comes, will impose a significant burden upon the agencies. They can imagine, in fact, that by virtue of their participation in DBWG they are in the position of assisting in the specification of a potentially heavy and unanticipated workload for their own agencies. Not

surprisingly, the members of DBWG, who are generally not in senior supervisory roles, feel uncomfortable in putting forward such specifications. They would feel much more comfortable if senior personnel in their agencies were brought up to date on DIDS developments and made aware that a White House request was on its way and what was the rationale for that request.

The members of DBWG appreciate the fact that the Federal Council on Statistical Plans and Programs is being convened to discuss the National Indicators System and that this body is likely to be apprised of the necessity for achieving an operational data base for DIDS in the very near future. This meeting should assist DBWG in the task of approaching the Economics and Statistics Service of USDA, the National Center for Health Statistics, the National Center for Education Statistics and the Energy Information Administration, agencies which seem to have no direct involvement in DIDS at present.

Many members of DBWG are from nonstatistical agencies, agencies that will nonetheless be asked to provide data. Unfortunately, a number of agencies from whom data will be requested are not participating in the DIDS program. Members of DBWG are encountering some refusals to cooperate--sometimes even from within their own agencies--when they ask for candidate lists of data elements for the operational data base. The refusals are encountered when agency personnel realize that these candidate lists will come back to haunt them in the form of a White House request. At the least, the agencies want to know what is going on before they supply the lists.

It is the consensus of DBWG that agencies need a communication from the DIDS program *as soon as possible.* Agencies need to know what is going on as regards the operational data base with special reference to the likely workload which the agencies will be asked to undertake.

Attachment 2 is a draft letter which might be sent by DIDS to appropriate agency officials in order to address the above concerns.

Attachments

1. Tentative List of Candidate Items
2. Draft Letter to Heads of Agencies

DRAFT

May 14, 1981

TENTATIVE LIST OF CANDIDATE ITEMS FOR DIDS OPERATIONAL DATA BASE

I. BUREAU OF THE CENSUS

1. 1980 PL94-171, Special Population Summary File	State, County, Tract
Total population	
Race	
Spanish origin	
2. 1979 County Business Patterns	State, County 4 digit SIC
No. of establishments	
Mid-March employment	
First quarter and annual payroll	
No. of establishments by employment size class	
3. 1978 Census of Agriculture (preliminary)	State, County
No. of farms	
Size of farms	
Value of farms	
Use of land	
Market value of agricultural products	
Farm organization and operator characteristics	
Livestock and poultry inventory	
Major crops	
4. 1978 General Revenue Sharing	State, County
Population estimates	
Per capita income	
5. 1977 Census of Government (employment)	State, County
Expenditures for corrections, education, administration, fire protection, health, libraries, police protection, public welfare, utilities, parks, transit, sewerage and highway maintenance.	
6. 1977 Census of Government (finance)	State, county
Revenue from taxes (amounts by tax source)	
Expenditures	
Debts	
7. 1980 Current Population Survey (March)	State
Household characteristics	
Income	
Age, sex, race	
Marital status	
Education	
Employment characteristics	
Insurance	
8. 1977 County and City Data Book	State, County
Population -1970,1975	
Vital statistics - 1975	

ETC., ETC., ETC.

[Substantially longer list will be provided by Tuesday, May 19, including many more agencies.]

DRAFT - What are your most frequently used databases?

May 14, 1981 Authoritative requests

TENTATIVE LIST OF CANDIDATE ITEMS FOR DIDS OPERATIONAL DATA BASE

I. BUREAU OF THE CENSUS

1. 1980 PL94-171, Special Population Summary File — State, County, Tract
   - Total population
   - Race
   - Spanish origin
   - [*Note*: Other data from 1980 Census to be acquired when available]
2. 1979 County Business Patterns — State, County 4 digit SIC
   - No. of establishments
   - Mid-March employment
   - First quarter and annual payroll
   - No. of establishments by employment size class
3. 1978 Census of Agriculture (preliminary) — State, County
   - No. of farms
   - Size of farms
   - Value of farms
   - Use of land
   - Market value of agricultural products
   - Farm organization and operator characteristics
   - Livestock and poultry inventory
   - Major crops
4. 1978 General Revenue Sharing — State, County
   - Population estimates
   - Per capita income
5. 1977 Census of Government (employment) — State, County
   - Expenditures for corrections, education, administration, fire protection, health, libraries, police protection, public welfare, utilities, parks, transit, sewerage and highway maintenance.
6. 1977 Census of Government (finance) — State, county
   - Revenue from taxes (amounts by tax source)
   - Expenditures
   - Debts
7. 1980 Current Population Survey (March) — State
   - Household characteristics
   - Income
   - Age, sex, race
   - Marital status
   - Education
   - Employment characteristics
   - Insurance
8. 1977 County and City Data Book — State, County
   - Population -1970,1975
   - Vital statistics - 1975
   - Employment - 1975
   - School enrollment (1975)

Health - 1975
Income - 1969, 1974
Family income - 1969
Public aid, social security benefits - 1976
Banking - 1976
Housing - 1970
Building permits - 1975- 1976
Government finance and employment - 1971, 1972
Elections - 1972, 1976
Crimes - 1975
Manufactures - 1972
Wholesale trade - 1972
Retail trade - 1972
Selected services - 1972
Mineral Industries 1972
Agriculture - 1974

9. 1974-1977 Consolidated County and City Data Book — State, County
   Subject matter same as 1977 County and City Data Book
10. 1975-2000 Projections of State Population, Age, Race, Sex (by State)
    Total population
    Race
    Age
    Sex
11. Current Population Surveys — State
    1979 farm wages
    1979 fertility
    1980 voting
12. 1979 State and Metro Area Data Book — State
    Election, President - 1969-1976
    Elections, U.S. Senators - 1970-1978
    Elections, U.S. Representatives - 1970-1978
    Voting-Age Population - 1978

II. BUREAU OF LABOR STATISTICS

[*Note*: Selection of years and other time periods to be included will be decided at later date.]

1. Industry Employment, Hours, and Earnings - State and Area
   Earnings, hourly, production workers — Industry by State,area
   Earnings, weekly production workers
   Employment
   Hrs of work, weekly, production workers
2. Insured Employment and Wages
   Contribution, UI employer — Industry by State
   Employment
   Reporting units, UI employer
   Wages, average annual
   Wages, average weekly
   Wages, total
   Wages, UI taxable
3. Unemployment and Labor Force - State and Area
   Labor force, civilian — State, SMSA, LMA, county, city, CETA programmatic area
   Employment
   Employment
   Unemployment

4. Consumer Price Index
   - Price index - item by city/U.S. average
   - Price index - item by region
   - Price index - item by city-size class
   - Price index - item by region/city-size class

III. SMALL BUSINESS ADMINISTRATION
(Call Fred Reikert)

IV. ENVIRONMENTAL DATA BASE - MULTI-AGENCY

1. Climate
   - Temperature
   - Average for each month, by county
   - Annual average by county
   - Average heating degree days by county
   - Average cooling degree days, by county
   - Average number of frost-free days by county
   - Precipitation
   - Average for each month, by county
   - Annual average, by county
   - Average annual snowfall, by county
2. Land cover and use
   - Cropland by county - 1950-1960-1970-1980
   - Forest, by county - 1980
   - Rangeland, by county - 1980
   - Urban development, by county - 1980
   - Ownership by state - % Federal, State, private, other - 1980
3. Cropland and agriculture
   - Major crops by State, 1980
   - Erosion by State, 1980
4. Critical and designated areas
   - Wilderness area, by state, 1980
   - Park area, by county, 1980
   - Coastline length, by State
   - Coastline ownership, by State, 1980
   - Area of wildlife refuge, by State, 1980
   - Designated historical sites, by county, 1980
   - Tornado frequency, by county
   - Seismic risk areas, by county
   - Hurricane risk areas, by county
5. Wildlife
   - Wildlife populations, by state, 1980
   - Gamefish catch, by state, 1980
   - Licensed fishery, by state, 1980
   - Licensed hunters, by state, 1980
   - Selected wildlife data from Geoecology
6. Water resources
   - Mean annual runoff by hydrologic accounting unit
   - Runoff by HAU, each year 1975-80, as % of mean annual
   - Current hydrologic conditions map, free form
   - Water use by State, 1975
7. Toxic substances
   - Production, by type, by county
   - Use of agricultural chemicals by State, 1980
   - Production of toxic waste, by state, 1980

8. Air quality
   - PSI's, each yr. 1975-80, days in ranges over 100 by SMSA
   - Criteria pollutants, violation rates, each yr 1975-80, by SMSA
   - Emissions, by type, by state, each yr. 1975-79
   - Compliance and noncompliance areas by SMSA, each yr 1975-80
9. Water quality
   - All NASQUAN constituents, 1975-80, by HAU
   - EPA region X index, by reach, 1979
   - Discharge permits, by reach, 1980
   - Identified contaminated ground water, by county, 1980
10. National Coal Resources Data System - USGS
11. Oil and Gas Production File - USGS
12. Water Use 5-year summary - USGS
13. Streamflow Current Conditions Map - USGS
14. Water quality by river basin - USGS
15. Pollution Discharge Inventory- USGS
16. Palmer Drought Severity Index by Climatic Zone - NOAA
17. Soil Moisture Index by Climatic Zone - NOAA

V. National Center for Education Statistics

1. Public School Universe — State
   - Names and address of school, county, district, programs offered, no. of pupils and teachers
   - 1969-72 4-yr. summary, 1977-78
2. Nonpublic School Universe
   - Name, mailing address, scholl district, curriculum, enrollment, no. of administrative staff, classroom teachers, total staff.
   - 1970-71; 1976-77; 1976-79
3. Merged Federal Files
   - Files include: NCES school district universe; Census of Govts. survey of local govt. finances, Office of Civil Rights elementary/secondary civil rights compliance survey; Office of Education Section 437 State-administered Federal programs; EEOC elementary/secondary staff survey; equalized property value compiled by Killalea Associations; and NIE special tabulations of 1970 census data for school districts
   - One-time, 1976-77
4. Opening Fall Enrollment
   - Covers full-time, part-time and first-time-equivalent enrollment, by sex, for undergraduate, graduate, unclassified, and first-professional degree students
   - 1966 through 1979, annual
5. Financial statistics
   - Includes current fund revenues, by source; expenditures, by function; physical plant assets, by type and value; indebtedness on physical plant; endownment revenues, changes in fund balance.
   - 1965-66 thru 1978-79
6. Faculty Salaries
   - Covers salary, tenure, and fringe benefits of full-time instructional faculty , by sex, length of contract, type of institutional unit.
   - 1971-72 thru 1979-80
7. Earned degrees conferred
   - Covers bachelor's, master's, doctor's and first professional degrees conferred in major disciplines and sub-fields; degrees and awards conferred for less than 4 yrs work beyond high schoool, by sex. 1975-76 and 1976-77 surveys contain racial/ethnic data within major disciplines.
   - 1968-69 thru 1978-79
8. Institutional characteristics
   - Includes institution name, address, telephone number, congressional district, type of control, level of offerings, types of programs, accreditation, names of principal officers, previous year's enrollment
   - 1969-70 thru 1979-80

9. HEGIS Longitudinal File
   HEGIS is Higher Education General Information Survey. All survey data from financial statistics form. Salaries by sex and length of contract for full-time instructional faculty, full-time, part-time, and full-time equivalent enrollment by sex. Plus misc. information on Carnegie and NCHEMS codes. Office of Education indication of loan delinquency and defaults, etc.
   1974-78 - 4 yrs.
10. Vocational Education Annual Report Data
   Financial status carryover funds and formula grant program information; performance report data, no. of teachers, status of teacher training and local administrative staff; enrollments in programs, by level and discipline code, and by special funded program.
   1974 thru 1977
11. Public Libraries (LIBGIS)
   Includes name and location of service outlets; population of area served; no. of volumes of books and serials stocked; circulation; no. of interlibrary and bulk loans; no. of positions filled by library staff; operating receipts and expenditures during fiscal year
   FY74; FY78
12. Public School Libraries
   Includes type and level of school; libraries and media centers; no. of employees, by level of education and full- or part-time statusp; print and non-print resources; expenditures

VI. U.S. Fire Administration - Federal Emergency Management Administration

1. Participation in National Fire Incident Reporting System -- % Fire Dept. per county
2. Fire Fatalities per Million - county averages for years 1974-1977
3. Fire Fatalities per Million - county averages for years 1974-1978

VII. Bureau of Economic Analysis

1. State personal income by type - quarterly, annual - 1968-1981
2. State per capita income - quarterly, annual
3. County personal income by type - annual - 1959, 1962, 1965 to date
4. County per capita income - annual

VIII. Economics and Statistics Service, Department of Agriculture

1. County Crop Estimates - 1972-1980 - annual data
2. Crops Data Base - by State
   Field crops - 1954-1980
   Potatoes - 1954-1980
   Seeds - 1963-1980
   Noncitrus fruits and nuts- 1963-1980
   Peppermint and spearmint - 1963-1980
   Wrinkled peas for seed - 1963-1980
   Citrus -1972-1980

Draft Letter to Heads of Agencies from Whom DIDS Will Request Data

Dear

The Decision Information Display System is viewed as a major tool in the National Indicators System, a program for systematically informing the President, Vice President and Cabinet Councils of the social, demographic and economic trends in America in a policy-relevant format. Initially developed under White House initiative as a demonstration project, DIDS is currently in the process of converting to an operational basis.

Under the direction of an interagency Steering Committee, the DIDS program has been studying for some months the question of an operational data base. The study has been carried out principally through an interagency committee known as the DIDS Data Base Working Group which has had participation from a broad spectrum of agencies. The aim of the working group has been to specify a set of data which should have first priority for installation into the DIDS system as its first operational data base. The group has looked chiefly toward those data which are in highest demand by users and are the firstline data products of the agencies. The group has examined only data which are already being released to the public (unless the agency for reasons of its own has chosen to make new data available).

Within the next several weeks, I expect that the Data Base Working Group will complete its task of specifying the first set of data for an operational DIDS data base. At that time, assuming that the DIDS Steering Committee ratifies the recommendations of the group, I will send letters to the agencies in my capacity as Chairman of the DIDS Steering Committee requesting that the recommended data be provided. The request will include the stipulations that the data be accompanied by on-line data documentation and keywording, that the data be in the DIDS exchange format, and that agencies adopt procedures for routinely updating and maintaining the data they provide.

The purpose of the present letter is to ask your cooperation with the Data Base Working Group in specifying first priority data resources from your agency and to provide you with the opportunity to begin planning for compliance with the DIDS data requests. The operational data base will be utilized in preparing briefings for the President, Vice President, Cabinet Councils and White House staff, as well as other senior Administration officials. I feel sure that you will recognize the importance of this undertaking and want to make sure your agency's data holdings are well represented in the DIDS data base.

Sincerely,

Dr. Richard S. Beal, Assistant to the President
Chairman, DIDS Steering Committee

TAB G

National Aeronautics and
Space Administration

Washington, D.C.
20546

AUG 4 REC'D

Attachment #7

file NASA

Reply to Attn of: ET

JUL 29 1980

Bear contact NASA.

Mr. Joseph W. Duncan
Co-Chairman
DIDS Executive Committee
Office of Federal Statistical
Policy and Standards
Department of Commerce
Washington, DC 20230

Dear Mr. Duncan:

Thank you for your letter of July 15, 1980, and your progress report on the Domestic Information Display System (DIDS). Naturally, I am very disappointed to hear that the intended transition to fully budgeted operational status has been delayed beyond Fiscal Year 1981.

As you know, NASA participation in DIDS has been as the source of technology and technical support, and not as a user of the system. Because of our role, and primarily because we are not an agency with a data base or strong needs for socio-economic data, NASA has not been a participant in funding the DIDS project.

The $75,000 provided in Fiscal Year 1980 and to which you referred in your letter, was provided specifically for the purchase of a DIDS terminal for the Council of State Governments (CSG). Normally those funds would have come from CSG, but in the interest of time, the NASA Technology Transfer Program provided them directly to the Goddard Space Flight Center to procure the terminal for CSG.

Given this background and, coincidentally, the heavy budget cuts for Fiscal Year 1981 in the Technology Transfer Program, NASA will not make a funding contribution for DIDS in Fiscal Year 1981.

Naturally, we will continue to do our best to provide technical support and assist in the beginning of your operational phase.

Sincerely,

Anthony J. Calio
Associate Administrator for
Space and Terrestrial Applications