OVERVIEW of the NASA ASTROPHYSICS DATA SYSTEM

a presentation to the

NASA NSI

USERS WORKING GROUP MEETING

R. B. POMPHREY

13 FEBRUARY 1991

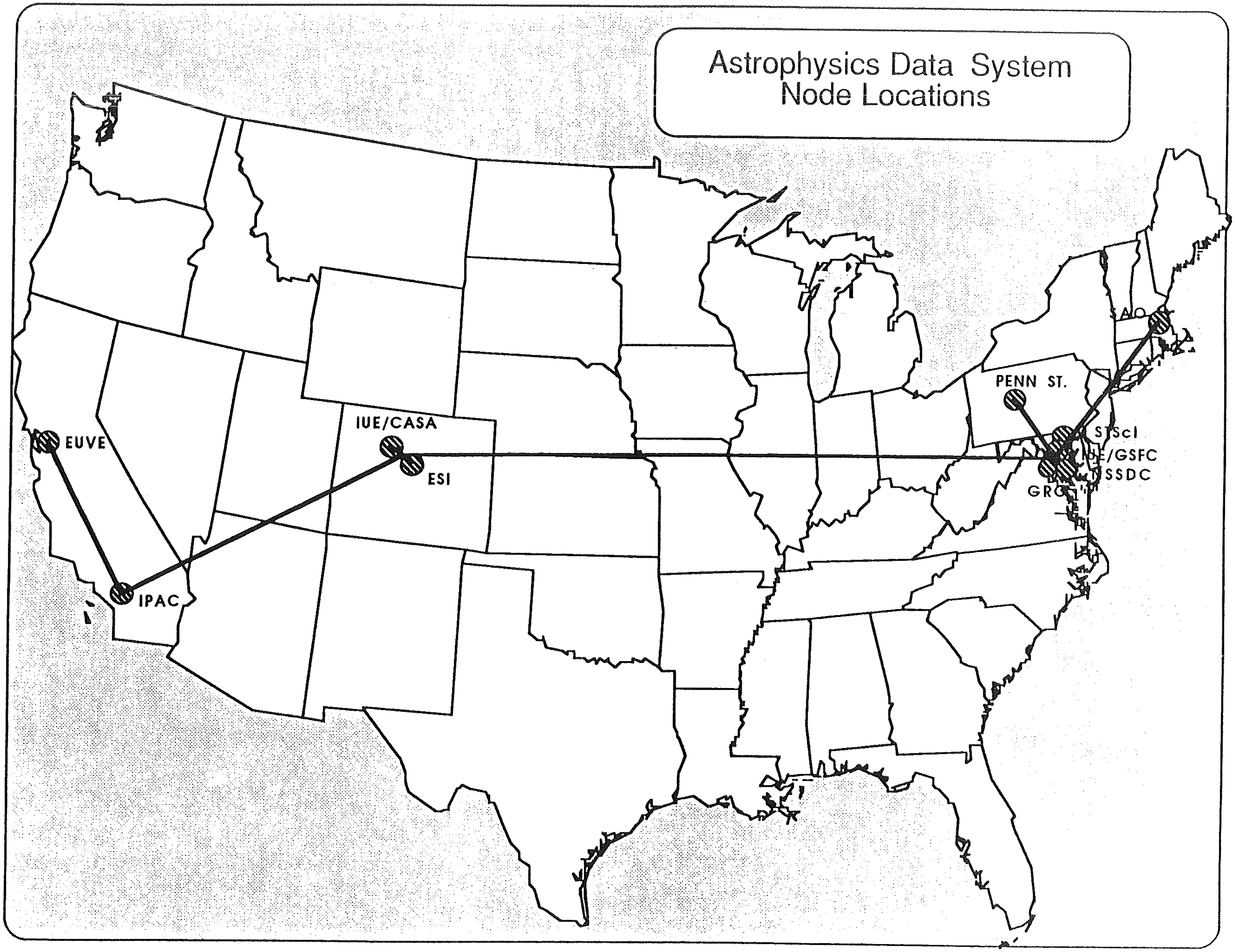
OVERVIEW of the ADS
AGENDA

- o The Problem
- o The ADS Project
- o Architectural Approaches
- o Elements of the Solution
- o Status of the Effort
- o The Future

OVERVIEW of the ADS

The Problem

- o Drinking from the Fire Hose
- o Multi-Spectral Research
- o Real-time Observing / Coordinated Observations
- o "Data about Data" and Tapping Human Expertise
- o Collaboration


Astrophysics Data System
Node Locations
EUVE
IPAC
IUE/CASA
ESI
PENN ST.
SAO
STScI

ADS OVERVIEW
THE ADS PROJECT

NASA ASTROPHYSICS DIVISION PROGRAM

DR. GUENTER RIEGLER -- MISSION OPERATIONS BRANCH CHIEF

DR. FRANK GIOVANE -- ADS PROGRAMS MANAGER

ASTROPHYSICS DATA SYSTEM PROJECT

DR. JOHN GOOD -- PROJECT MANAGER

DR. STEPHEN MURRAY -- PROJECT SCIENTIEST

DR. JOHN NOUSEK -- USER COMMITTEE CHAIR

ELLERY SYSTEMS, INC. -- SYSTEMS INTEGRATION

ADS OVERVIEW
THE ADS PROJECT - continued

CHARTER

To Provide current and future generations of space scientists with direct, on-line access to existing and future multispectral data and analysis tools.

OBJECTIVE

The ADS is a production level distributed processing system. The Objective of the ADS is to make all science data holdings and all ADS Hardware and Software services available to all users transparently.

OVERVIEW of the ADS

Architectural Approaches

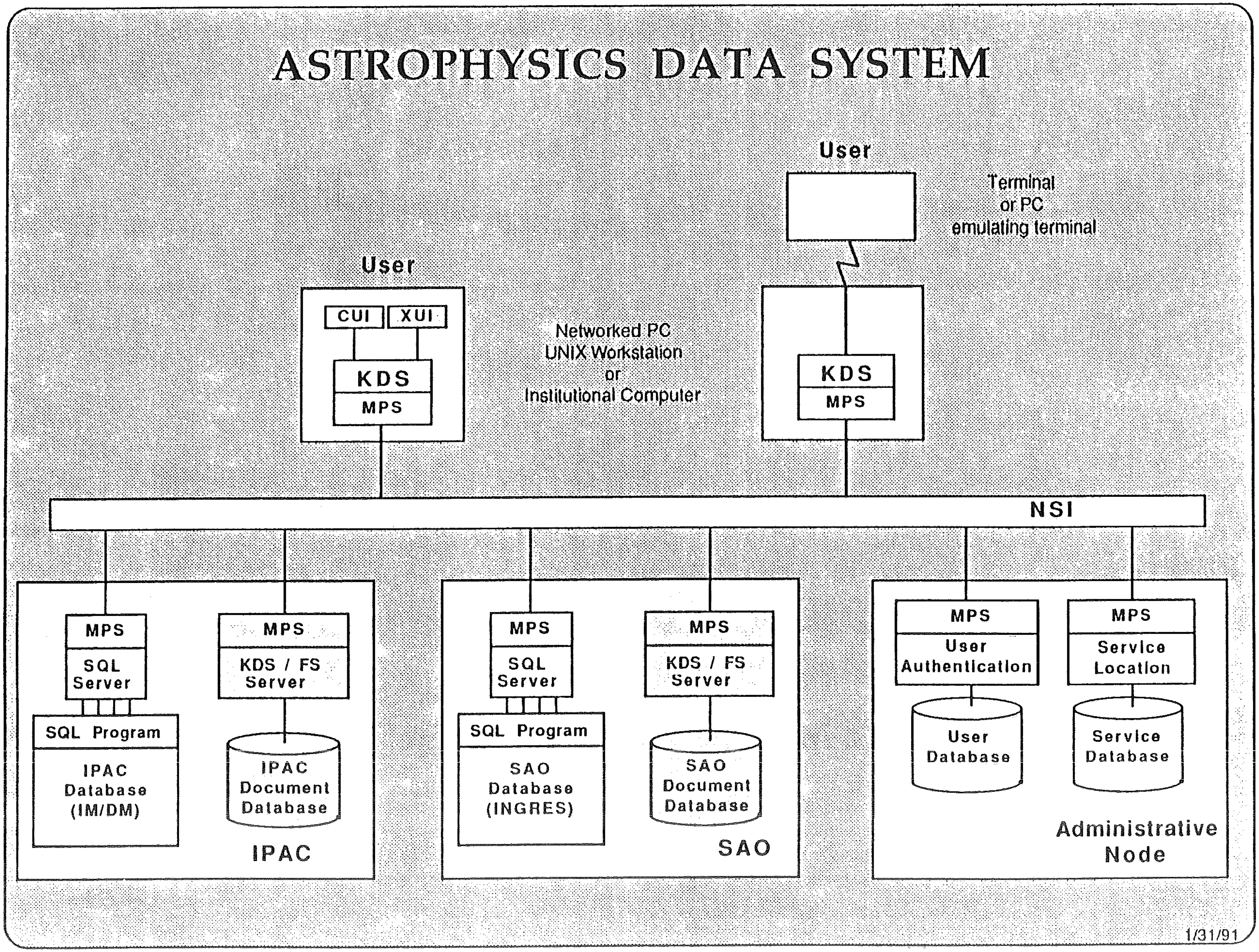
- o The "rlogin" model
  - User accesses each site independently
  - User must have accounts everywhere
  - Tools and interfaces are generally site specific
  - Data Transfer is done in a different environment

- o The Client / Server Model
  - Global Uniformity
  - Standard Interfaces
  - Modularity
  - Separation of Processing and User Interface
  - Easily incorporates existing services
  - Easily expanded and evolved
  - Location independence (of user, data, and processing tools)

OVERVIEW of the ADS

Elements of the Solution

- o User Interfaces
- o User Services
  - Distributed Access to Existing Database Systems
  - Document Location and Retrieval
  - Local Table Manipulation
  - Local Visualization
- o System Services
  - User / Service Authentication
  - Help
- o Glue
  - NASA Science Internet
  - Message Passing Service
  - Standard Data Formats and I/O


ASTROPHYSICS DATA SYSTEM
User
Terminal or PC emulating terminal
User
CUI
XUI
KDS
MPS
Networked PC
UNIX Workstation
or
Institutional Computer
KDS
MPS
NSI
MPS
SQL Server
SQL Program
IPAC Database (IM/DM)
MPS
KDS / FS Server
IPAC Document Database
IPAC
MPS
SQL Server
SQL Program
SAO Database (INGRES)
MPS
KDS / FS Server
SAO Document Database
SAO
MPS
User Authentication
User Database
MPS
Service Location
Service Database
Administrative Node
1/31/91

ADS OVERVIEW

ELEMENTS OF THE SOLUTION - SQL SERVER

- o Heterogeneous DBMS's
  - Relational
  - Heirarchical
  - Network
- o Distributed Interaction
- o Homogenization and Translation

ADS OVERVIEW

ELEMENTS OF THE SOLUTION - FACTOR SPACE

- o Failure of the "Library" Model
- o Personal Perspective
- o Fields and Terms
- o Factor Spaces

ADS OVERVIEW

DEVELOPMENT AND OPERATIONS PHILOSOPHY

The ADS was designed and built by practicing astrophysicists for practicing astrophysicists.

Utilizing the most advanced commercially available and supported distributed processing system technology, it is specifically designed to meet the evolving needs of the professional scientist and to provide the community with a powerful and immediately useful research and educational facility.

ADS OVERVIEW

STATUS

PHASE ONE:

At present, data holdings from SAO and IPAC are accessible using advanced remote procedure call and other advanced distributed processing system techniques. Over the next three months, data holdings from IUE/GSFC, IUE/CASA, STScI, Penn State, and the NSSDC will be added to the system. Data holdings from all great observatory and explorer class missions will be added as available.

PHASE TWO:

Provide on-line access to existing and future data analysis and manipulation tools to include imaging/visualization, graphic analysis, statistical, and such other tools as deemed appropriate by the user community. These tools will be made available as distributed processed to maximize compute and software resource availability to the community.

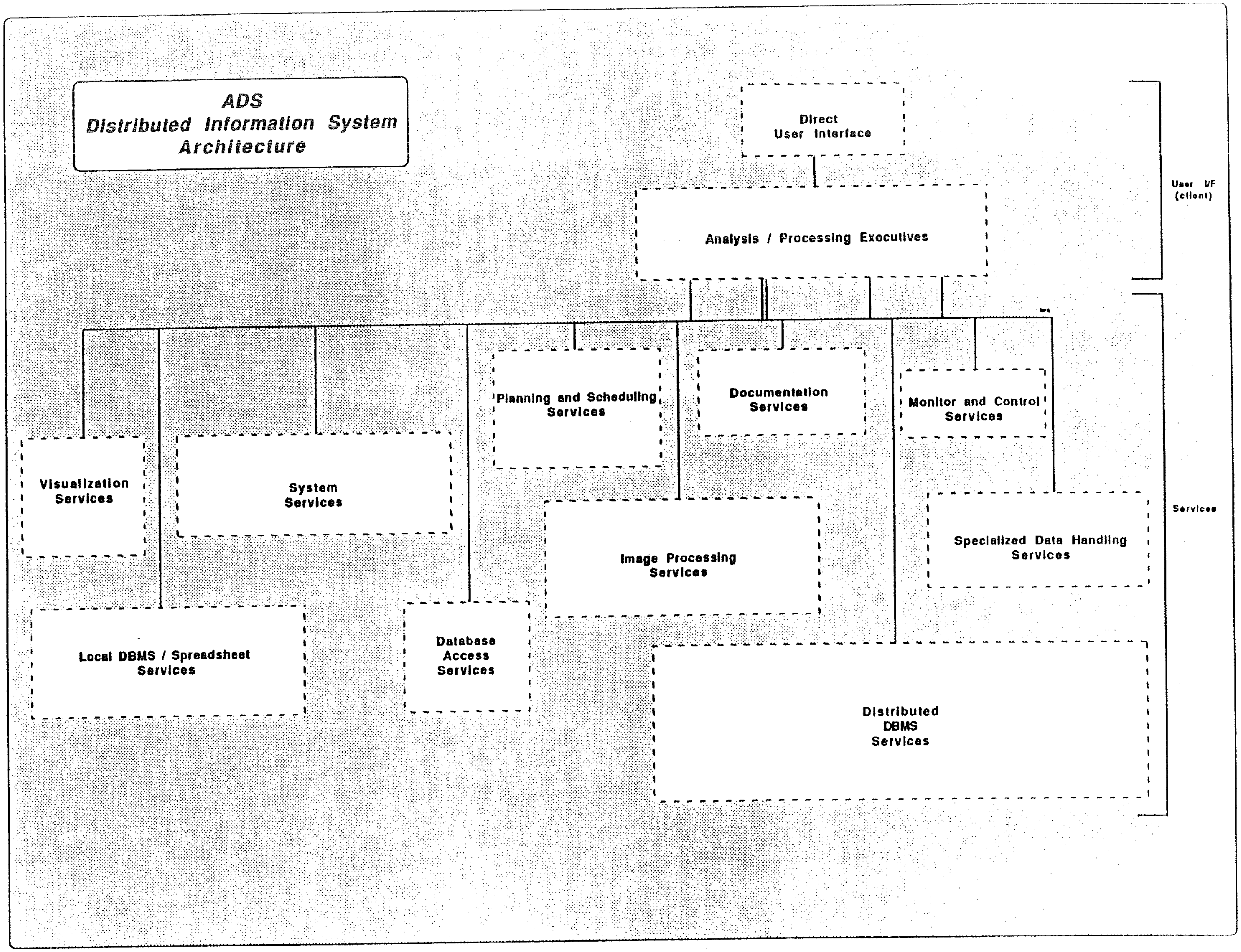
ADS OVERVIEW
THE FUTURE

- o System Generated Services
  - Transaction Management
  - System Monitoring
  - System Interfaces
  - User Interfaces
  - Communications Services

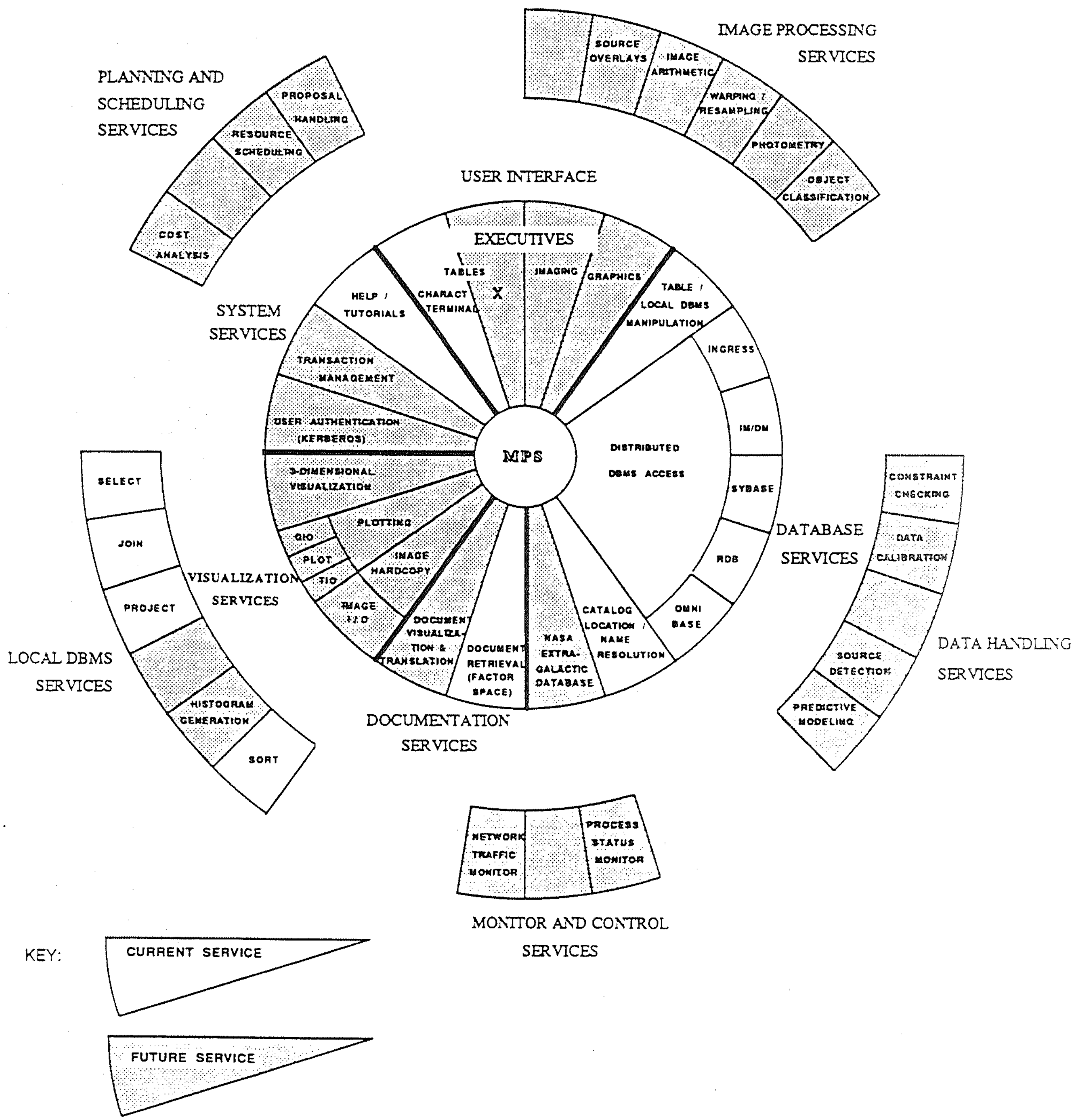
- o User Generated Services
  - Data Analysis
  - Visualization
  - Modeling

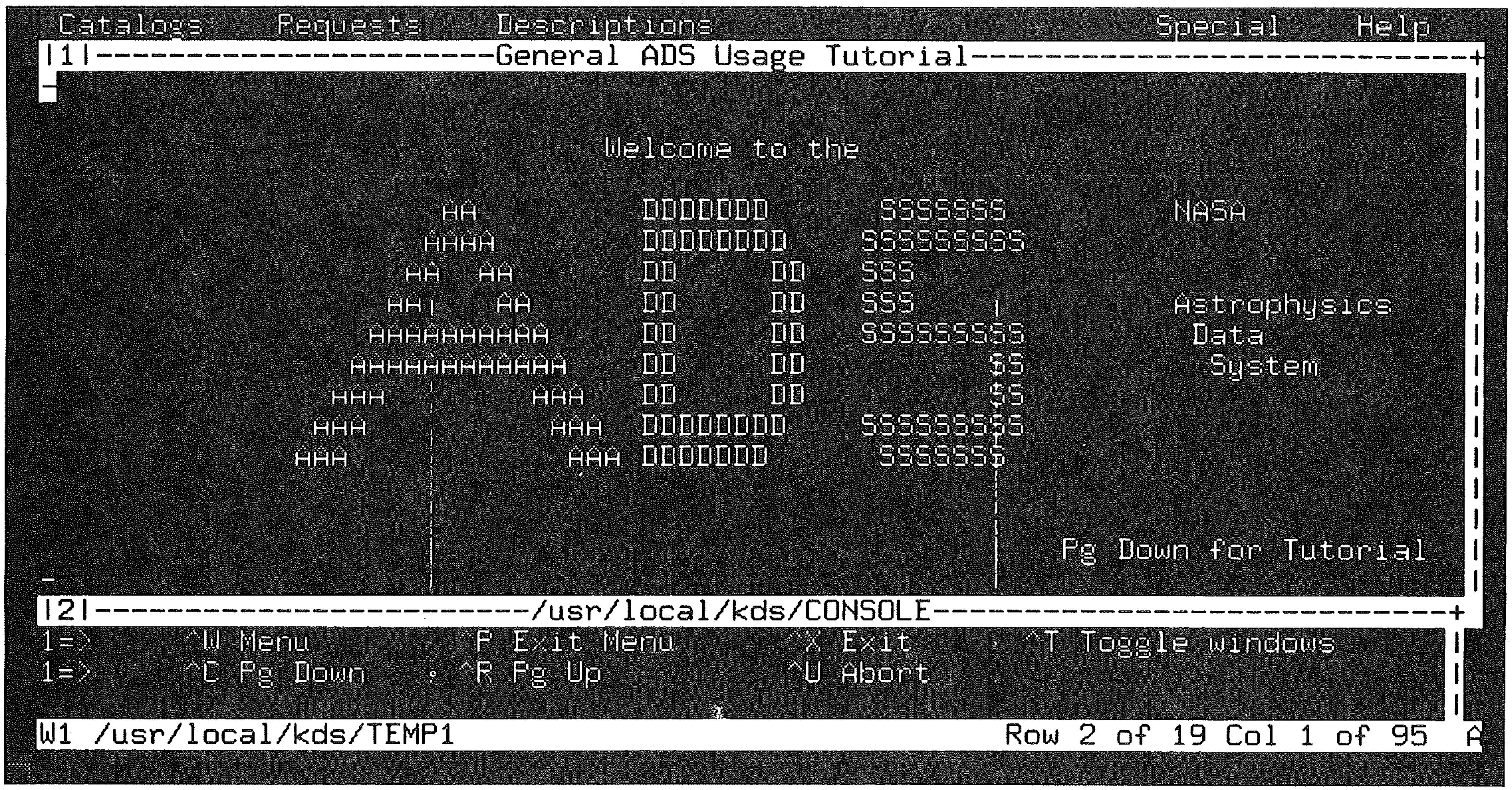
ADS OVERVIEW
THE FUTURE - continued

- o Project Generated Services
  - Planning and Scheduling
  - Monitor and Control
- o Datasets
  - Images
  - Spectra
  - Ground - Based Data
  - Textbooks and Journals


ADS
Distributed Information System Architecture
Direct User Interface
Analysis / Processing Executives
User I/F (client)
Visualization Services
System Services
Planning and Scheduling Services
Documentation Services
Monitor and Control Services
Image Processing Services
Specialized Data Handling Services
Local DBMS / Spreadsheet Services
Database Access Services
Distributed DBMS Services
Services

ASTROPHYSICS DATA SYSTEM
FUNCTIONAL ARCHITECTURE



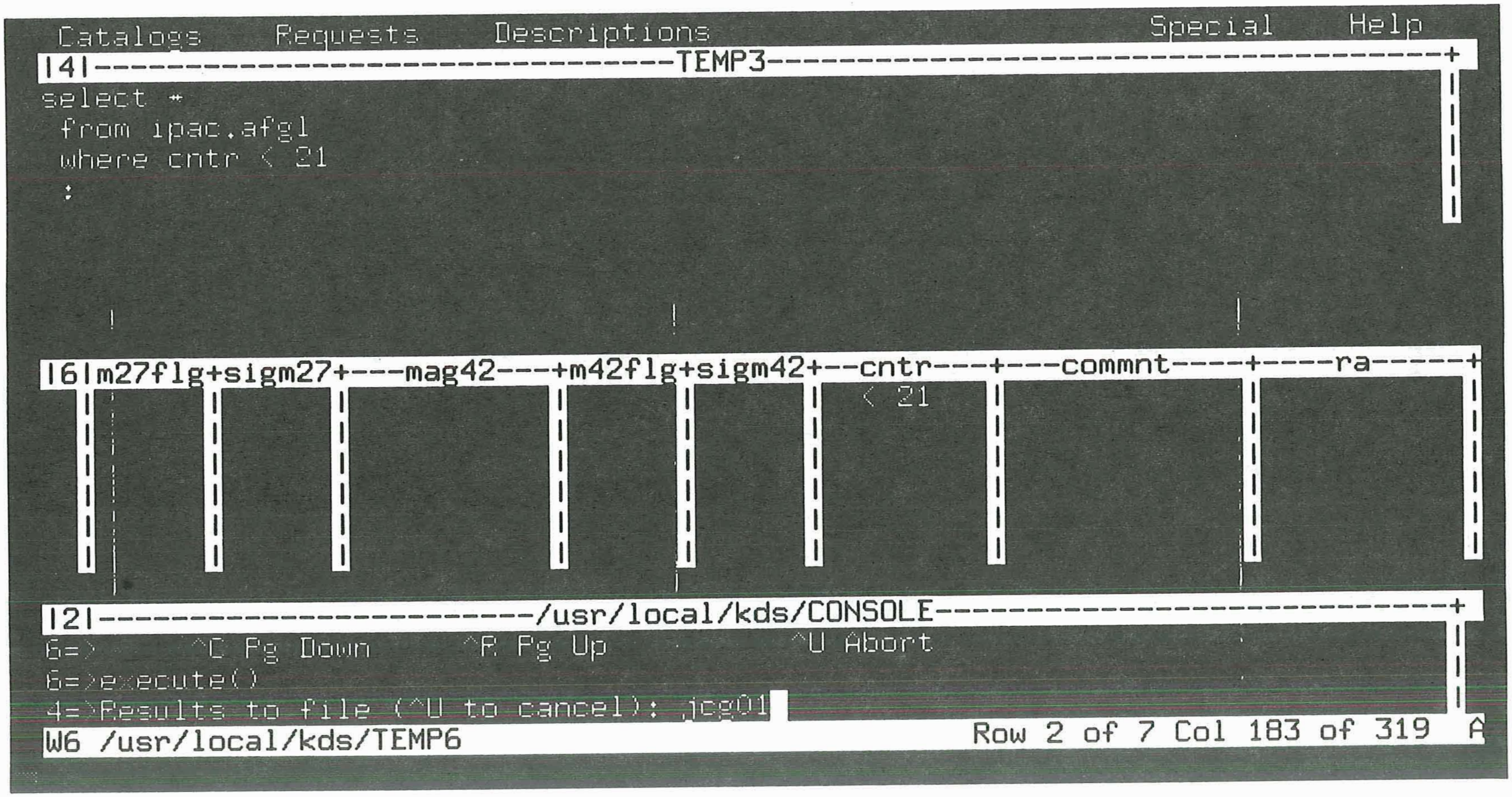

Catalogs Requests Descriptions Special Help
|1|----General ADS Usage Tutorial----+
Welcome to the
AA DDDDDDD SSSSSSS NASA
Astrophysics
Data
System
Pg Down for Tutorial
|2|----/usr/local/kds/CONSOLE----+
1=> ^W Menu ^P Exit Menu ^X Exit ^T Toggle windows
1=> ^C Pg Down ^R Pg Up ^U Abort
W1 /usr/local/kds/TEMP1 Row 2 of 19 Col 1 of 95

```
 Catalogs    Requests    Descriptions                          Special    Help
|3|Name--+------------Astronomical Catalog Name-----------------------------------
 eopsc   |   Einstein Observing Log
 irpsc   |   IRAS Point Source Catalog Version 2.1
 afgl    |   Revised Air Force Geophysics Lab Infrared Sky Survey
 bsc     |   Bright Star Catalog, 4th Edition
 cgcg    |   Catalog of Galaxies and Clusters of Galaxies
 cio     |   Catalog of Infrared Observations
 drbn    |   Dearborn Obs Cat of Faint Red Stars
 eic     |   Equatqrial Infrared Catalog
 eso     |   ESO/Uppsala Survey of the ESO (B) Atlas
|4|---------------------------ADS_INSTRUCTIONS-----------------------------------+
                                                                  <listselect>  |
                  Cursor to row.  Enter ^E                                      |
                                                                                |
Move the cursor to the desired item.  The window will scroll if there          |
is too many entries to be displayed at once.  With the cursor on the           |
correct line, enter ^E .  The line in the window will be highlighted           |
and the selection will be complete.                                            |
|2|-----------------------/usr/local/kds/CONSOLE---------------------------------+
3=>        ^W Menu        ^P Exit Menu        ^X Exit        ^T Toggle windows  |
3=>        ^C Pg Down     ^R Pg Up            ^U Abort                          |
                                                                                |
W3 /usr/local/kds/TEMP3                             Row 4 of 24 Col 1 of 95   A
```

```
 Catalogs    Requests    Descriptions                             Special     Help
|5|------------The Revi+--------------------+ Survey Catalog-------------------
<Description>          |.                  . |
The "Revised AFGL Infra|Catalog desc         |g" (RAFGL, Price and Murdock 1983)
contains the results of|Field list           |on and photometric studies of AFGL
sources performed since|.                  . |he AFGL Four Color Infrared Sky
Survey:  Catalog of Obs|(select field)       |0, 9.8, and 27.4 microns (AFGL,
Price and Walker 1976) +--------------------+vey measurements with larger
instruments.

Stephan D. Price
Thomas L. Murdock
16 June 1983
AFGL-TR-83-0161

<TOC>

   Annotations:  a. Some fields/elements are INDEXED to provide quicker access
                    to data values. These fields/elements are annotated as "Gen"
|2|-------------------------/usr/local/kds/CONSOLE--------------------------------+
5=>       ^W Menu          ^P Exit Menu        ^X Exit        ^T Toggle windows
5=>       ^C Pg Down       ^R Pg Up            ^U Abort
5=>^FO
W5 /usr/local/kds/TEMP5                                Row 5 of 22 Col 14 of 95
```

```
 Catalogs    Requests    Descriptions                          Special    Help
|4|Name--+----Units-----+------------Description-----------------------------+
 M11FLG  |              |   '<' - mag 11 is an upper limit                   |
 M20FLG  |              |   '<' - mag 20 is an upper limit                   |
 M27FLG  |              |   '<' - mag 27 is an upper limit                   |
 M42FLG  |              |   '<' - mag 4.2 is an upper limit                  |
 MAG11   |   Mag        |  Magnitude of 11 microns.  '-9.9' - unavailable    |
 MAG20   |   Mag        |  Magnitude of 20 microns.  '-9.9' - unavailable    |
 MAG27   |   Mag        |  Magnitude of 27 microns.  '-9.9' - unavailable    |
 MAG42   |   Mag        |  Magnitude of 4.2 microns. '-9.9' - unavailable    |
 OBSLOG  |              |  Observation log                                   |
|5|------------The Revised AFGL Infrared Sky Survey Catalog-------------------
2.  COLUMNS <MAG42> <MAG11> <MAG20> <MAG 27> <SIGM42> <SIGM11> <SIGM20>
      <SIGM27> <M42FLG> <M11FLG> <M20FLG> <M27FLG>

    The 4-, 11-, 20-, and 27-micron photometry is listed in the next four
columns, respectively, along with the estimated error or source reference if
it is not a survey measurement.  A magnitude derived from the CIO listing is
designated by a C, one taken from Grasdalen et al^A15^ by a W and a value
|2|----------------------/usr/local/kds/CONSOLE-----------------------------+
4=>        ^W Menu        ^P Exit Menu        ^X Exit        ^T Toggle windows |
4=>        ^C Pg Down     ^R Pg Up            ^U Abort                         |
W4 /usr/local/kds/TEMP3                           Row 13 of 30 Col 1 of 95   A
```

```
Catalogs    Requests    Descriptions                              Special    Help
|6|m27flg+sigm27+---mag42---+m42flg+sigm42+--cntr---+---commnt----+----ra-----+
                                            < 21

|7|---------------------------ADS_INSTRUCTIONS--------------------------------+
                                                                     <qbe>
    p. to print.  (value) for = value   Enter ^E to execute

A complete dissertation on the usage of QBE is beyond the scope of this
help (see the ADS User's Guide).  Basically, control sequences are placed
in each column to determine the SQL used for a DBMS search.  The basic
sequences are:
|2|----------------------/usr/local/kds/CONSOLE-------------------------------+
6=>     ^W Menu        ^P Exit Menu       ^X Exit        ^T Toggle windows
6=>     ^C Pg Down     ^R Pg Up           ^U Abort

W6 /usr/local/kds/TEMP6                           Row 2 of 7 Col 183 of 319  A
```


Catalogs Requests Descriptions Special Help
|4|---TEMP3---+
select *
from ipac.afgl
where cntr < 21
;
|6|m27flg+sigm27+---mag42---+m42flg+sigm42+--cntr---+---commnt----+----ra-----+
< 21
|2|---/usr/local/kds/CONSOLE---+
6=> ^C Pg Down ^R Pg Up ^U Abort
6=>execute()
4=>Results to file (^U to cancel): jcg01
W6 /usr/local/kds/TEMP6 Row 2 of 7 Col 183 of 319 A

Catalogs Requests Descriptions Special Help

7 GLNO	BSID	CDEC	CNTR	COMMNT	CRA	DEC	ELAT
25		+73 45 06	1		00 00 01.0	73.7517	61.73
35		+24 37 13	2		00 00 15.0	24.62	22.44
45		+58 17 30	3		00 00 20.0	58.2917	51.27
40025		+59 27 36	4		00 00 31.0	59.46	52.15
5		+55 24 24	5		00 00 44.0	55.4067	48.98
7	9095	+66 26 02	6		00 01 16.1	66.4339	57.14
40035		+64 52 30	7		00 01 40.2	64.875	56.03
8		+39 49 53	8		00 01 45.6	39.8314	35.82
9	9105	+41 50 43	9		00 01 56.0	41.845	37.55
60015		-01 46 40	10	EC	00 01 59.0	-1.77778	-1.828
60025		-02 09 10	11	EC	00 02 08.7	-2.15278	-2.188
60035		-01 43 32	12	EC	00 02 10.0	-1.72556	-1.798
5001		-01 51 25	13	EC	00 02 26.9	-1.85694	-1.947
5002		-02 08 33	14	EC	00 02 35.5	-2.14223	-2.223
60045		-02 07 50	15	EC	00 02 58.3	-2.13056	-2.250
60055		-43 15 44	16		00 03 02.2	-43.2622	-39.24
40055		+56 03 24	17		00 03 30.0	56.0567	49.26

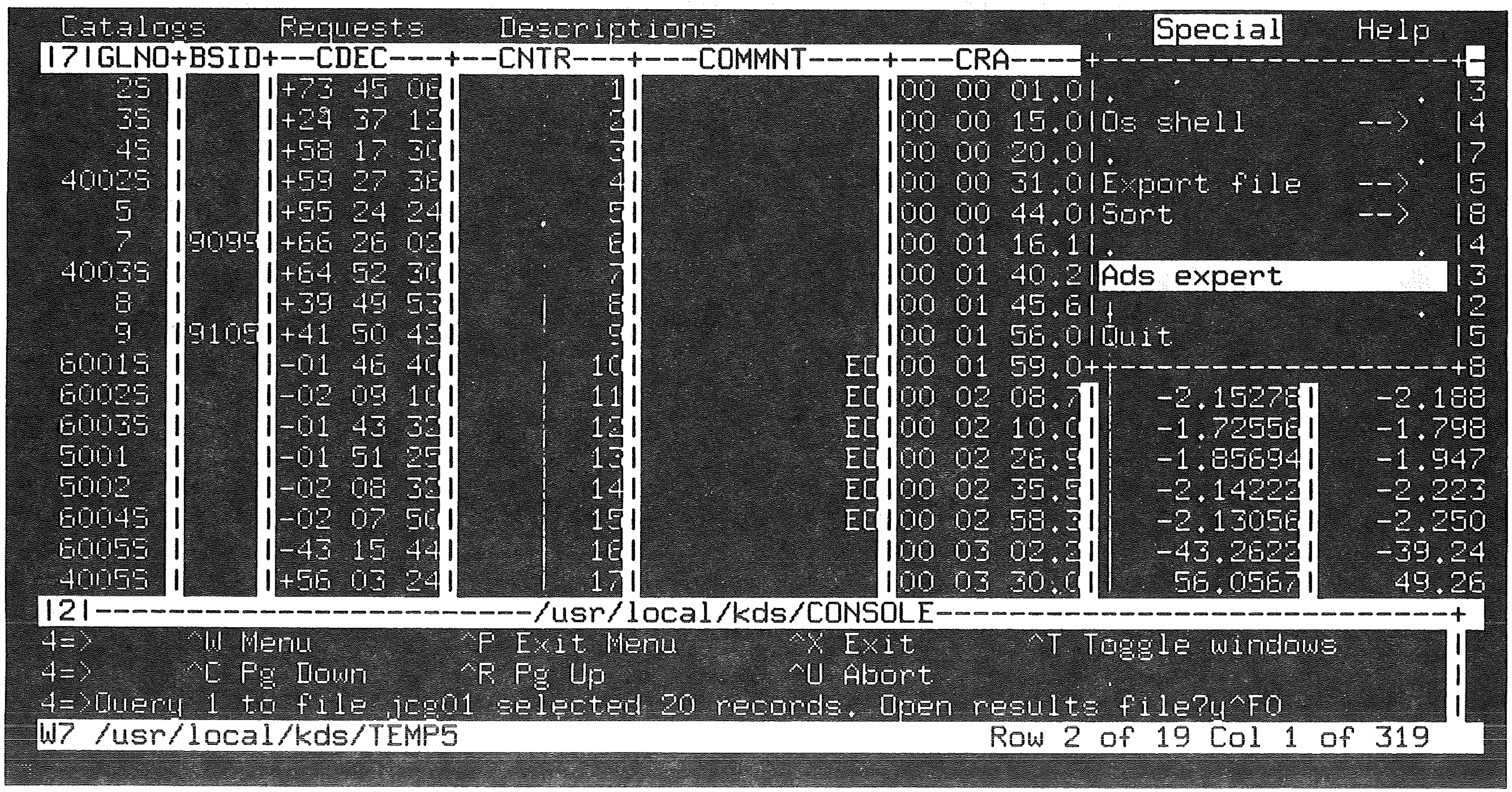
|2|------------------------/usr/local/kds/CONSOLE------------------------+

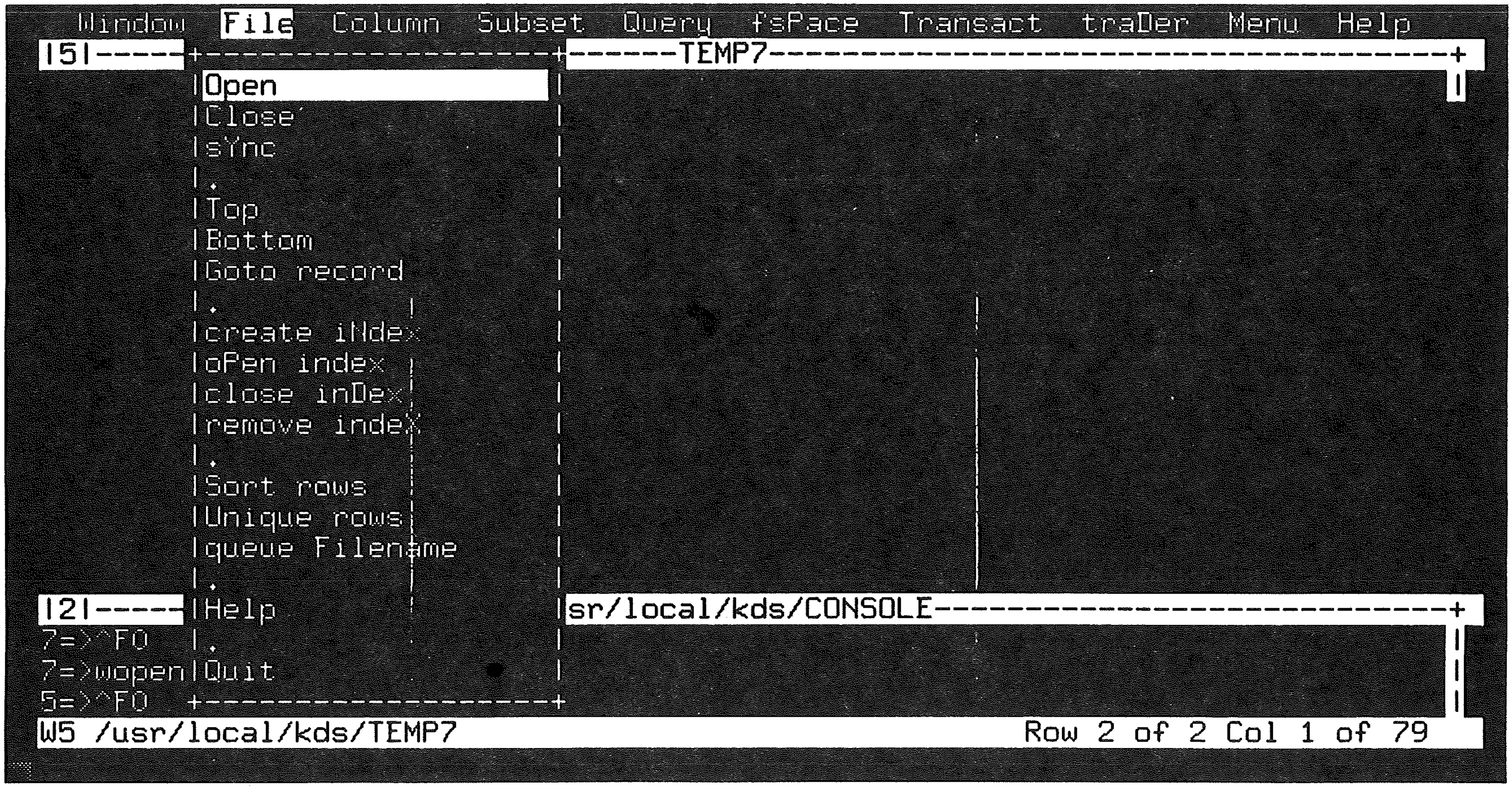
4=> ^W Menu ^P Exit Menu ^X Exit ^T Toggle windows
4=> ^C Pg Down ^R Pg Up ^U Abort
4=>Query 1 to file jcg01 selected 20 records. Open results file?y

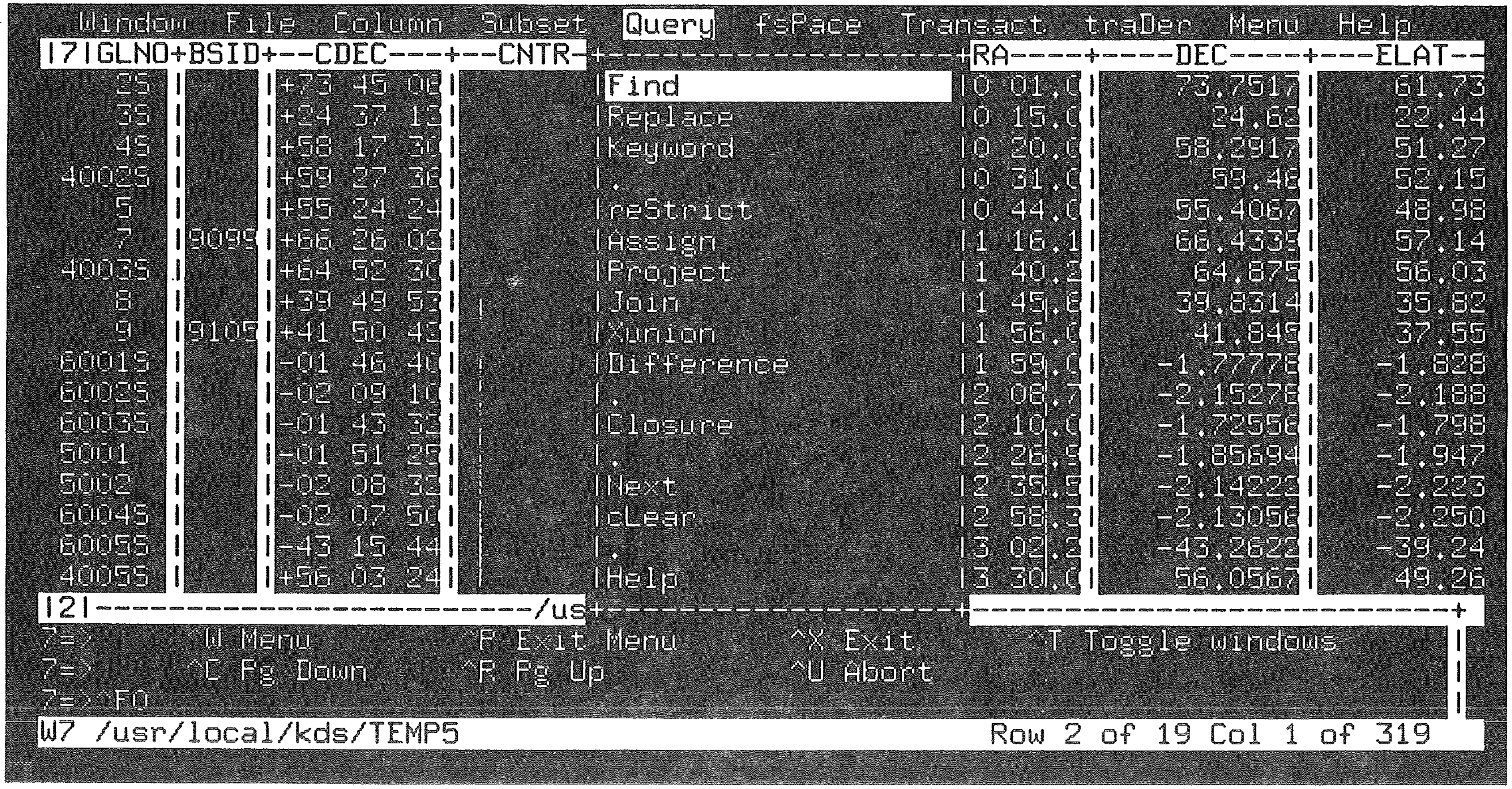
W7 /usr/local/kds/TEMP5 Row 2 of 19 Col 1 of 319 A

214

ORIGINAL PAGE IS OF POOR QUALITY


Catalogs Requests Descriptions Special Help
|7|GLNO+BSID+--CDEC---+--CNTR---+---COMMNT----+---CRA----+
25 | | +73 45 0 | 1 | |00 00 01.0|. . |3
35 | | +24 37 1 | 2 | |00 00 15.0|Os shell --> |4
45 | | +58 17 3 | 3 | |00 00 20.0|. . |7
40025 | | +59 27 3 | 4 | |00 00 31.0|Export file --> |5
5 | | +55 24 24 | 5 | |00 00 44.0|Sort --> |8
7 |9095 | +66 26 0 | 6 | |00 01 16.1|. . |4
40035 | | +64 52 3 | 7 | |00 01 40.2|Ads expert |3
8 | | +39 49 5 | 8 | |00 01 45.6|. . |2
9 |9105 | +41 50 4 | 9 | |00 01 56.0|Quit |5
60015 | | -01 46 4 | 10 | EC |00 01 59.0+
60025 | | -02 09 1 | 11 | EC |00 02 08.7 | -2.1527 | -2.188
60035 | | -01 43 3 | 12 | EC |00 02 10.0 | -1.7255 | -1.798
5001 | | -01 51 2 | 13 | EC |00 02 26.9 | -1.85694 | -1.947
5002 | | -02 08 3 | 14 | EC |00 02 35.5 | -2.1422 | -2.223
60045 | | -02 07 5 | 15 | EC |00 02 58.3 | -2.1305 | -2.250
60055 | | -43 15 44 | 16 | |00 03 02.3 | -43.2622 | -39.24
40055 | | +56 03 24 | 17 | |00 03 30.0 | 56.0567 | 49.26
|2|----------------------/usr/local/kds/CONSOLE----------------------------+
4=> ^W Menu ^P Exit Menu ^X Exit ^T Toggle windows
4=> ^C Pg Down ^R Pg Up ^U Abort
4=>Query 1 to file jcg01 selected 20 records. Open results file?y^FO
W7 /usr/local/kds/TEMP5 Row 2 of 19 Col 1 of 319


Window File Column Subset Query fsPace Transact traDer Menu Help
|5|-----
TEMP7
Open
Close
sYnc
Top
Bottom
Goto record
create iNdex
oPen index
close inDex
remove indeX
Sort rows
Unique rows
queue Filename
Help
Quit
|2|-----
sr/local/kds/CONSOLE
7=>^FO
7=>wopen
5=>^FO
W5 /usr/local/kds/TEMP7
Row 2 of 2 Col 1 of 79


Window File Column Subset Query fsPace Transact traDer Menu Help
|7|GLNO+BSID+--CDEC---+--CNTR-+--------------------+RA----+----DEC----+---ELAT--
25 | |+73 45 08| |Find |0 01.0| 73.7517| 61.73
3S | |+24 37 13| |Replace |0 15.0| 24.62| 22.44
4S | |+58 17 30| |Keyword |0 20.0| 58.2917| 51.27
4002S | |+59 27 38| |. |0 31.0| 59.48| 52.15
5 | |+55 24 24| |reStrict |0 44.0| 55.4067| 48.98
7 |9099|+66 26 03| |Assign |1 16.1| 66.4339| 57.14
4003S | |+64 52 30| |Project |1 40.2| 64.875| 56.03
8 | |+39 49 53| |Join |1 45.8| 39.8314| 35.82
9 |9105|+41 50 43| |Xunion |1 56.0| 41.845| 37.55
6001S | |-01 46 40| |Difference |1 59.0| -1.77778| -1.828
6002S | |-02 09 10| |. |2 08.7| -2.15278| -2.188
6003S | |-01 43 33| |Closure |2 10.0| -1.72558| -1.798
5001 | |-01 51 25| |. |2 26.9| -1.85694| -1.947
5002 | |-02 08 32| |Next |2 35.5| -2.14222| -2.223
6004S | |-02 07 50| |cLear |2 58.3| -2.13058| -2.250
6005S | |-43 15 44| |. |3 02.2| -43.2622| -39.24
4005S | |+56 03 24| |Help |3 30.0| 56.0567| 49.26
|2|------------------------/us+--------------------+---------------------------+
7=> ^W Menu ^P Exit Menu ^X Exit ^T Toggle windows
7=> ^C Pg Down ^R Pg Up ^U Abort
7=>^FO
W7 /usr/local/kds/TEMP5 Row 2 of 19 Col 1 of 319

OVERVIEW OF THE ASTROPHYSICS DATA SYSTEM

John C. Good and Richard B. Pomphrey
Infrared Processing and Analysis Center
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91125

Abstract

The Astrophysics Division of NASA has built a geographically- and logically-distributed heterogeneous information system for the dissemination and coordinated multispectral analysis of data from astrophysics missions. The Astrophysics Data System (ADS) is a truly distributed system in which the data and the required processing are physically distributed. To accommodate the anticipated growth and changes in both requirements and technology, the ADS employs a server/client architecture which allows services and data to be added or replaced without having to change the basic architecture or interfaces. Current datasets accessible through the system include all the tabular astronomical data available at each of six existing astrophysics data centers. Additional data nodes, at both NASA data centers and academic institutions, will be added shortly. The future evolution of the system will be driven in large part by user services mounted both by the ADS project itself and by members of the astrophysics community.

Astrophysics Data System Philosophy and Strategy

Astrophysicists today have a bewildering array of powerful NASA resources to call upon. Among these are the data centers for the High Energy Astrophysics Observatories (HEAO's), the International Ultraviolet Explorer (IUE), the Infrared Processing and Analysis Center (IPAC), and the Hubble Space Telescope Science Institute (STScI), as well as the National Space Science Data Center (NSSDC). Unfortunately, the services provided by these centers are essentially independent and some are only accessible through mission-unique hardware and software. Furthermore, they can generally only be accessed directly through the centers themselves.

The Final Report of the Astrophysics Data System Study, March 1988, characterized the data environment of the Astrophysics community at that time and defined for the future an "... architecture for a data system which will serve the astrophysics community in multi-spectral research through the decade of the 1990's." One of its recommendations was that each data set, and the human expertise which supports the data, be maintained at the same physical location. Moreover, these multiple locations should be linked together and to researchers by means of high speed communications networks. Finally, to allow true multi-spectral research the various data sets should be accessible through a common set of tools.

The Astrophysics Data System (ADS) Project is NASA's response to the Data System Study. For the science investigator, the ADS makes NASA's current and future data holdings more broadly and efficiently accessible, and makes the data itself more interpretable. For NASA, it provides a common information system infrastructure for science analysis, thereby reducing duplication of effort while increasing the scientific return from missions.

The ADS is a truly distributed system in which both the data and the required processing are physically distributed. To accommodate anticipated growth and changes in both requirements and technology, the ADS employs a server/client architecture which allows services and data to be added or replaced without having to change the basic architecture or interfaces. The ADS design is modular and layered, enabling smooth evolution of the hardware and software without interruption of service to the user community. In addition, the ADS will provide all on-line information necessary to use the ADS services and data; and its design enables remote updating of the software services.

ADS Software Architecture

Conceptually, the software can be divided into two categories. Core Services are those which provide the basic system functionality upon which user applications or User Services can be built. These will be built primarily by the ADS Project itself. User Services are other applications which reside on the ADS and provide science support and analysis functions required by the investigator. These may be built by the Project in response to community demand or by individuals or groups funded by the NASA Astrophysics Science Research Aids (ASRA) Program.

In its initial release, the ADS will provide basic Core and User Services. These will be expanded and supplemented in the future based on feedback from the Astrophysics community.

Initial Core Services

The MESSAGE PASSING SERVICE (MPS) enables remote inter-operability and data transfer/translation among the ADS services. It provides homogeneity across networks and operating systems for process requests and responses. While this service is largely invisible to the end user, it provides the application programmer (and the ADS system programmers) what the User Interface provides for the science end user: an environment in which to access the services of the ADS from within an application program without knowing which servers will execute the program or the services it accesses. At a minimum, the MPS implements the functionality of:

- Remote Procedure Call (RPC): a mechanism by which the subroutines of a program can call each other as if they were executing on a single server when, in fact, they may be segmented across several servers;

- Remote Process Invocation (RPI): a mechanism by which a program on one server can spawn programs on other servers;
- Remote Inter-Process Communication (RIPC): a mechanism whereby two or more programs running concurrently on different servers can communicate among themselves.

In the initial release, the ADS Message Passing Service conforms to the Advanced Network System Architecture Testbench (ANSAT) and the Open Software Foundation/Distributed Computing Environment (OSF/DCE).

USER INTERFACES are the services which provide the local working environment (including its look and feel) to the end user. As such they provide a means to access all other services within the ADS. The structure of the ADS is such that there can be any number of user interface programs incorporating the constraints, hardware limitations, and preference of specialized segments of the community. Our efforts to date have concentrated on a single user interface which is structured to conform to a "least common denominator" hardware environment, and which provides an effective interface to the initial set of ADS Services. The ground rules have been:

- the ADS system must be accessible using any character terminal supportable under the UNIX "termcap" facility (basically any display with 24x80 characters and an addressable cursor);
- the ADS system must be effective (if slow) at 1200 baud.
- it must be possible (but not required) to configure the system such that the user interface and appropriate interactive processing is done at the user's site, rather than at centralized facilities.
- the user interface in particular should be available under UNIX, VMS and MS/DOS.

Even with these constraints, we are able to provide an ADS User Interface that provides the following combined functionality:

- specialized table display and interactive manipulation facilities;
- a first-order complete relational database facility including support for Structured Query Language (SQL) and Query by Example (QBE);
- Menu bar/pulldown menus and multiple windows (split-screen);
- context-sensitive help and a dynamic tutorial facility;
- full-featured text management facility that supports browsing, plain-text inquiry and cut-paste editing of selected files.

Through this interface, investigators can locate data sets (descriptions and catalogs) of interest, access these data sets either individually or in combination, and select from among them one or more subsets that they wish to import into a local working environment for further analysis of their own choosing. More importantly, it allows the investigator to accomplish this without requiring him to know which servers execute the services he invokes. In the initial release, the Knowledge Dictionary System (KDS) (tm) is being used to implement the functionality of the User Interface.

The DATA INDEPENDENT ACCESS MODEL (DIAM) is a service which provides inter-operability and language conversion among the various distributed Database Management Systems (DBMS's) in the ADS. These DBMS's include the following: Ingres and Sybase (UNIX), RDB (VMS), and IM/DM (CDC NOS/VE). The DIAM is required in the ADS in order to provide the user with a uniform relational view of all these distributed catalogs even though some of the catalogs are still maintained using DBMS's that do not yet support the relational SQL standard. The functionality of the DIAM is:

- to accept SQL statements generated by the User Interface (either directly or indirectly using the QBE translator);
- to convert them to the language supported by the DBMS that hosts the referenced catalog(s);
- to RPI a process to submit them to the DBMS for processing and to collocate the resulting table(s);
- to return those results to the User Interface that issued the request.

The Distributed Access View Integrated Database (DAVID) system is being used to implement the functionality of the DIAM.

Besides providing uniform access to the multiplicity of existing DBMS's, DAVID provides a complete internal distributed DBMS which allows further processing with special features not available in most commercial DBMS's. Of particular importance in a distributed environment is its ability to allow dataset browsing. This minimizes the overhead potentially incurred by having to transfer data in large chunks around the country.

Future Core Services

There are several Core Services which are important and anticipated in the near future, but which were not included in the initial release of the Astrophysics Data System. Among these are the User/Service Authentication Service and the Transaction Management Service. Other Core Services will be added on as they are required and become available.

The USER/SERVICE AUTHENTICATION SERVICE will provide the capability to automatically verify the authorization of a user to access the ADS System and to access specific services and data within the system. It will be provided through an implementation of the KERBEROS software on the ADS System.

The TRANSACTION MANAGEMENT SERVICE (TMS) will provide the process and resource management protocol for client-defined transactions. It assures successful execution, synchronization, and release of all services and resources used in a transaction. The TMS provides two basic functions:

- insure, without further client intervention, that all the services requested during a transaction will be successfully executed even if some of those services or the servers on which they are executing fail while the transaction is in progress;
- insure, without further client intervention, that all resources involved in a transaction are properly synchronized and released regardless of the destiny of the transaction (success or failure).

The TMS is both an optional and a passive service; optional in that it must be explicitly invoked by the client; and passive in that, as a peer-to-peer system, there is no mechanism by which the TMS protocol can be enforced, and the only programs that are guaranteed to participate in the TM protocol are the core services of the ADS. To encourage the use of the TM by end-users and application programmers, the protocol has been kept as simple as possible, requiring only four commands: Begin, Lock, Upgrade, and End.

Begin is a signal that a program is starting and returns a unique Transaction ID that the program must use in all subsequent calls to the TM. Lock signals the intent of the program to access the resource (e.g., a record in a file) named in the call. Upgrade signals the intent of the program to modify a previously Locked resource. End signals that the program is terminating, and the mode of termination (success or failure). For the initial ADS release, the components of the Transaction Management Service are implemented by the Transaction Manager (tm), which is an integral part of the Knowledge Dictionary System (tm). The TMS components will be exported as discrete services through the Message Passing Service through which they will be accessible by Remote Procedure Calls.

Initial User Services

The initial User Services available through the ADS have usually been collectively referred to as Directory Services, with components that provide access to catalog data and to documentation about data holdings for specific Astrophysics archives, without requiring the user to know where the data physically resides. These Directory Services include Document Retrieval, Documentation Browse, and a Catalog Data Retrieval and Processing Service. Astrophysics data comes in several forms (e.g., catalog data, spectral data, image data). The initial release of the ADS will be limited to catalog data (though some of the catalogs are in fact lists of images or spectral observations).

The DOCUMENT RETRIEVAL SERVICE provides uniform, subject-matter indexing (and English-language querying) across all the data in the ADS, regardless of whether data are highly structured in databases, or unstructured. For the initial release, the Document Retrieval Service is implemented by the Factor Space Access Method (tm) as an integral part of the Knowledge Dictionary System (tm). The various components of the Document Retrieval Service will be exported as discrete services through the Message Passing Service, through which they can be generally accessed by Remote Procedure Calls.

The Factor Space (FS) is an n-dimensional Euclidian space, the axes of which are statistically constructed to account for the variance and covariance in expert judgments made by astrophysicists about the relevance of ADS data items to different subject matter contexts. The functions of the Document Retrieval service are:

- to scan all data entered in the ADS and compute its subject matter profile as a sequence of one or more vectors in the Factor Space;
- to similarly analyze natural language requests and to search the Factor Space for relevant data items;
- to monitor the distribution of vectors in Factor Space for clusters (i.e., undifferentiated data items);
- to periodically generate, disseminate, collect and synthesize questionnaires to obtain additional relevance judgments needed to increase the data resolution;
- to factor-analyze these additional relevance judgments and modify the number and/or orientation of axes in the Factor Space accordingly.

These functions also support the generation of new Factor Spaces, either to accommodate new subject matter or to accommodate personalized perspectives on existing subject matter.

The DATA BROWSE SERVICE provides a simple means for users to access and organize directories and files through the functionality of the User Interface described above.

The DATA RETRIEVAL AND PROCESSING SERVICE provides the capability to retrieve cataloged data and perform data base management processing on that data through the query, text management, and relational data base functionality of the DIAM and User Interface services described above.

Future User Services

It is expected that other User Services will be made available as they are requested by the Astrophysics user community and integrated into the ADS. Because of the flexible modular nature of the ADS architecture, such integrations should be relatively straightforward and can be implemented in a variety of ways, dependent on how the prospective User Service is coded. These new services can be derived from any of the following sources:

- Astrophysics Science Research Aids (ASRA) Program
- NASA Astrophysics Flight Projects
- Other NASA Flight Projects
- Other NASA Programs
- Non-NASA Programs
- Non-US Programs

ADS Physical Architecture

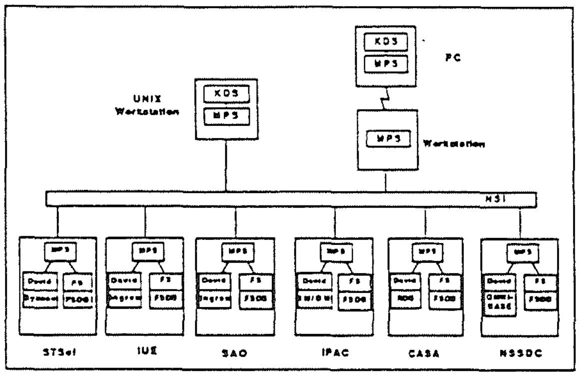
An overview of the ADS physical architecture is presented in the figure below. In its initial instantiation, the ADS will consist of six physically distributed primary nodes which are interconnected via NASA Science Internet (NSI). These six nodes are the following:

- Center for Astronomy and Space Astrophysics CASA), Boulder, Colorado
- Infrared Processing and Analysis Center (IPAC), Pasadena, California
- International Ultraviolet Explorer (IUE), Greenbelt, Maryland
- National Space Science Data Center (NSSDC), Greenbelt, Maryland

- Smithsonian Astrophysical Observatory (SAO), Boston, Massachusetts
- Space Telescope Science Institute (STScI), Baltimore, Maryland

Platforms and Operating Systems

Each of these nodes above has important astrophysics catalogs for which it has principal responsibility. In general, these catalogs are maintained in different types of DBMS's on different types of machines, each of which runs a unique operating system. Components of the ADS Core Services will be resident on a server at each node and connected by the ADS Message Passing Service software. (See the ADS Primary Nodes Compatibility Requirements Document for a more detailed description of the hardware interface.) The mapping of services to servers in the ADS is unconstrained: some servers simultaneously provide several (simple) services while some (complex) services are segmented across several servers. In addition, the numbers and kinds of services that can be mounted on the ADS are also unconstrained.



Physical/Logical Organization of Representative ADS Services.

Nodes and Sites

The ADS will be composed of Primary and Secondary Nodes, an Administrative Node, and User Sites.

An ADS primary node is defined as a facility which assumes primary responsibility for the provision of a unique service through the ADS infrastructure. This service can involve the provision of basic mission-specific data, as would be the case for a mission science support center like IUE, or for the provision of an operational service for remote access such as the IPAC-supported NASA Extragalactic Database (NED).

A Primary Node assumes the responsibility for user support, maintenance of user services, data, and appropriate documentation it provides via the ADS. Many different kinds of services may be offered by primary nodes in the future, such as access to data archives or use of specialized processors or software.

A Secondary Node is not responsible for user support or for maintenance of data or service provided by the node. Its role is to provide local or regional copies of data or services. Provision for such copies should alleviate the load on the primary nodes and greatly improve the responsiveness and reliability of the system.

The ADS will also maintain an administrative node whose primary functions are to monitor the system (the user/service database, network throughput and connectivity, usage patterns, service availability, and security), and to serve as the top of the hierarchy that deals with user questions, problems and training.

A site is defined as any physical location, outside of an ADS Node, where ADS-specific software exists (e.g., the User Interface, MPS, or DIAM) through which an investigator accesses the ADS or its services.

Networking

The first release of the ADS required network connectivity among the primary nodes and remote users (those not physically located at one of the primary nodes). This network connectivity will be furnished by the NASA Science Internet (NSI) which supports the TCP/IP protocol. While it is intended that the ADS will eventually support the DECNET protocol, this will not be the case for the first ADS release.

User Scenario

The prospective user of the Astrophysics Data System should first obtain a copy of the ADS User's Guide (contact the administrative node at the address at the end of this document), which will give specific details of how to obtain access to the ADS System. Generally a user will be assigned to one of the ADS Primary Nodes for user support.

It is assumed that the user is a novice and is trying to do a very simple, well-defined task but one which requires access to multiple astrophysics data centers and their on-line databases. In this scenario, the user will locate and subset two independent datasets and then intercompare the results. Specifically, the problem is to correlate measurements of galaxies which have been observed to have both a significant infrared flux (indicative of a large amount of cold dust), and an X-ray or ultraviolet flux (indicative of hot gas).

The user first asks about the existence of 100 μm (infrared) data, and ultimately finds his/her way to the IRAS Point Source Catalog. This user then wants to know if a subset of IRAS sources (e.g. galaxies that lie above 30 degrees galactic latitude with 100 μm flux greater than 10 Janskys and 60/100 μm flux ratios indicative of temperatures less than 40K) have been observed in the X-ray (Einstein Catalog).

Further, for those sources which were observed both in the infrared and with Einstein list the ones detected in IR and X-ray and plot the IR flux versus the X-ray flux for these objects.

The specific steps are outlined below:

Locating Infrared Datasets

If we consider all the datasets potentially available in astrophysics (not just the 'standard' products of NASA missions), the potential size and diversity of these datasets become quite large. Without data location tools, locating the correct dataset for our investigation would be at least as difficult, if not more so, than the processing we will do once the data is found. In most cases accessing this data would be much more difficult than processing it.

One of the primary services of the initial ADS implementation is the factor space documentation location method described above. With this we can pose a factor space query of the type "I am interested in long wavelength infrared measurements of galaxies, particularly 100 μm measurements, in the form of catalogs. Color temperatures, specifically ones derived from long wavelength measurements, will be correlated with X-ray data to try and determine the relationship between cold dust emission and that from hot gas".

The factor space query will return a list of documents, some of which might be journal articles about similar research, some descriptions of catalogs or other data (e.g., images), some mission descriptions, etc. Somewhere near the top of this list is likely to be the description of the IRAS Point Source Catalog. Unless we have been distracted by some of the other documents, we will read this and decide that this catalog is indeed what we want for our IR data. By poking around the documentation related to this catalog (a simple browse mechanism through our documentation) we also determine what fields in the catalog are important for our investigation.

The documents used in the search are maintained at the same institutions that bear responsibility for the data itself.

Locating the X-ray Datasets

The X-ray data (in particular the Einstein database) would be found the same way. One of the rules of the ADS is that catalog datasets are not made accessible through the system until appropriate documentation is also online.

Extracting the IR Data

We have determined the correct catalog and even have some idea as to which fields are appropriate for our endeavor. The next step is to actually query the database where the data resides and get our subset. The basis for such queries is SQL, as described in the section on our DIAM (DAVID). Our query will look something like:

```
select * from iraspsc where flux100>10 and
  flux60<flux100 and (glat>30 or glat<-30)
```

In addition to a direct SQL query capability, ADS provides a more intuitive query by example (QBE) mechanism where the user puts constraint information directly into a template of the catalog in question.

The infrared data of interest is located at the Infrared Processing and Analysis Center at the California Institute of Technology in Pasadena, California on a CDC Cyber mainframe in the IM/DM database.

Extracting the relevant X-ray Data

To compare sources which both have an X-ray flux in the Einstein catalog and correspond to the sources obtained from our infrared query, requires the use of SQL for a database join. This is complicated somewhat by the need to potentially transform coordinates from one system or representation and then to perform the join based on spherical distance proximity. This process can also be performed using SQL directly or through the QBE mechanism described above.

Selection of the data must be done at the site where the data resides. However, the join process can be done at either data site or at the user's local site. Optimization can be done to try and minimize the amount of data transfer but in practice the user usually wants to see the results of the two 'selects' before joining them and therefore will transfer the data to the local site anyway.

The X-ray data will be at the Harvard-Smithsonian Astrophysical Observatory in Cambridge, Massachusetts in INGRES on a SUN workstation.

Comparison of IR and X-ray Data

Once the join has been completed, the result is a single table in which are combined IR and X-ray measurements. Proximity on the sky was the deciding factor in matching sources and so we may well want to edit this table in a spreadsheet manner before proceeding. The final step in this example is simply to plot one column in this table versus another.

Table processing and plotting will be performed locally to the user.

System Management Summary

The ADS Program is currently managed from the Science Operations Branch, Astrophysics Division, NASA Headquarters, Washington, D.C. under Guenter Riegler. The ADS Project has been established at the Infrared Processing and Analysis Center which has also been designated as the ADS Administrative Node. Dr. John Good is the ADS Project Manager and is responsible for the activities of the Administrative Node.

Management of the overall ADS effort includes System Development, System Oversight, and System Administration. What follows is a summary of the more detailed material documented in the ADS Management Plan.

System Development

The design of the ADS is motivated by the fact that it exists in a dynamic science and information system environment, and therefore it must be a dynamic system. Thus, even as the ADS is released, development continues on Core Services, User Services, Network Connectivity, and the addition of new nodes. System Management is responsible for technical oversight of prototyping research, selection criteria, development of standards and conventions, and all aspects of systems integration, test, and operation of new or revised ADS services.

System Oversight and Review Process

To date, oversight of ADS development has been provided by the ADS Working Group, under the leadership of James Weiss and John Nousek. When ADS becomes operational (and new nodes and sites are added to the system), the operation and continued development of the ADS will be overseen from both a users' and a systems perspective by an anticipated hierarchy of committees. These committees will review all proposals for ADS development, and will integrate, prioritize, and make policy recommendations on all aspects of the ADS Program.

The Science Operation/Management Operations Working Group (SOMOWG) has the highest oversight and review responsibility. As a matter of policy, the Science Operations Branch will make final selection of new ADS services, based on the results of the proposal and peer review process. For this purpose, an annual NASA Research Announcement (NRA) for the Astrophysics Software and Research Aids (ASRA) Program will be issued.

System Administration

The administration of the Astrophysics Data System is shared by NASA Headquarters, Astrophysics Division (Programmatic), the ADS Project Office/ADS Administrative Node (Project), and the other primary nodes making up the ADS. The organizational structure, and defined roles and responsibilities are documented in Appendix A and the text of the ADS Management Plan, and in the ADS Primary Node Compatibility Requirements.

Development responsibilities include administration of both in-house and external service development contracts, certification of software, and software licensing.

Operations Responsibilities include administration of maintenance contracts, system monitoring, system change control, and maintenance of system documentation.

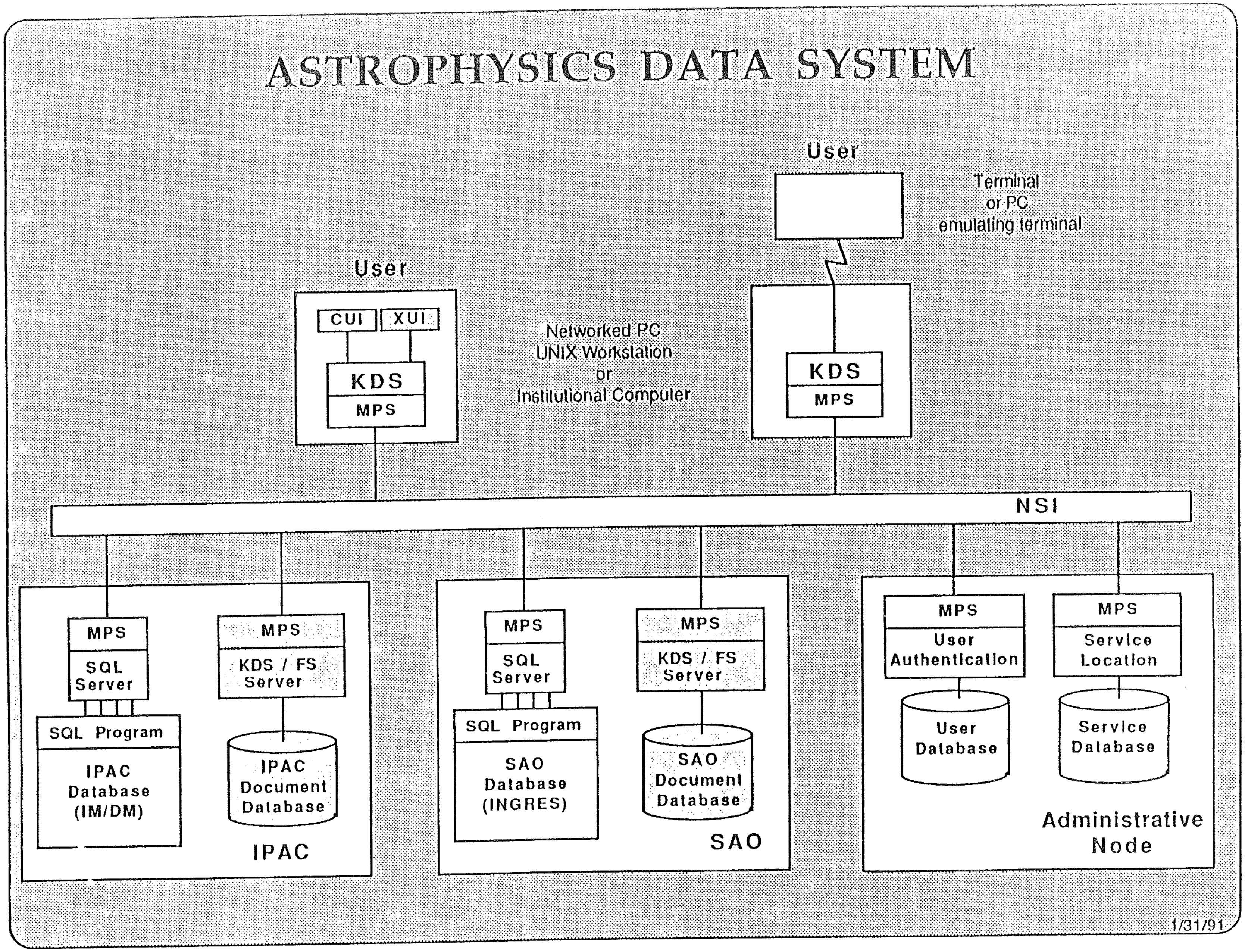
In addition, the administration and coordination of all aspects of the review processes relevant to the ADS are the responsibility of the System Administration.

Further information on the ADS can be obtained by contacting Dr. Good at

Internet	jcg@ipac.caltech.edu
BITnet	jcg%ipac@HAMLET.BitNet
Telemail	[JGOOD/NASA]NASAMAIL/USA
uucp	(cit-vax,trwrb!csula-ps)!ipac!jcg
SPAN	ROMEO::"jcg%ipac"

TEL: (818) 584-2939
FAX: (818) 584-9945

The work described in this paper was supported by the Jet Propulsion Laboratory, California Institute of Technology, under the sponsorship of the Astrophysics Division of NASA's Office of Space Science and Applications.


ASTROPHYSICS DATA SYSTEM
User
Terminal
or PC
emulating terminal
User
CUI
XUI
KDS
MPS
Networked PC
UNIX Workstation
or
Institutional Computer
KDS
MPS
NSI
MPS
SQL
Server
SQL Program
IPAC
Database
(IM/DM)
MPS
KDS / FS
Server
IPAC
Document
Database
IPAC
MPS
SQL
Server
SQL Program
SAO
Database
(INGRES)
MPS
KDS / FS
Server
SAO
Document
Database
SAO
MPS
User
Authentication
User
Database
MPS
Service
Location
Service
Database
Administrative
Node
1/31/91