10

The NASA/IPAC Extragalactic Database

G. Helou, B.F. Madore, M. Schmitz, X. Wu, H.G. Corwin, Jr., C. LaGue, J. Bennett, & H. Sun

10–1 introduction

The NASA/IPAC Extragalactic Database (NED) is an electronic research tool which provides uniquely powerful access to published multi-wavelength data on extragalactic objects. It consists of a computer database with a broad range of published information, and a user interface which allows fast and flexible retrieval of the information via the Internet. It is accessible anonymously and free of charge to any account on the networks, and used primarily by researchers, students and librarians. NED is physically located at the Infrared Processing and Analysis Center (IPAC) on the Caltech campus in Pasadena, California (USA).

NED provides an object-based representation of the extragalactic sky, with accurately located sources and coherently presented data, and direct links from this information to the refereed literature. NED is built around astronomical sources. Its starting point is a detailed merger of extragalactic catalogs, rather than a collection of juxtaposed catalogs. This merger is constantly enriched by new identifications and by the addition of objects appearing in short lists throughout the literature. The resulting database is augmented with a systematic bibliography, with well-characterized measurements for each object, and with notes from catalogs and the literature. NED can be searched for objects in a variety of ways, including constraints on positions, redshifts and type of object, and will plot the distribution on the sky of objects retrieved. NED can be searched for references by citation or by author name, and will display abstracts of papers published since 1988, and of theses from English-language institutions.

In the four years since its public release, NED has been adopted by the extragalactic research community worldwide. The response from users has been extraordinary, both in the enthusiasm expressed and the frequency of

Published in Information & On-line Dafa in Astronomy,
Editors: D. Egret & M. A. Albrecht, Kluwer Academic Publishers, Dordrecht, 1995.

consultation. As of the Fall of 1994, this frequency hovered around 6,000 sessions per month, with hundreds to thousands of additional batch-mode requests and server-mode queries per month.

10–2 Context

Lists of galaxies were being assembled as part of the early surveys for diffuse nebulae (Messier, Heschel, etc.) even before the term "extragalactic" had been coined. The advent of photographic sky surveys led to the first catalogs of galaxies with tens of thousands of entries (*e.g.*, the CGCG by Zwicky *et al.*). The number of galaxies in catalogs has grown exponentially with time, from about thirty at the time of Messier to the millions being generated today by plate-scanning machines like APM (Cambridge University) or APS (University of Minnesota). At the same time, specialized catalogs of galaxies (interacting, low surface brightness, de.) and of quasars have also multiplied.

Over the past three decades, advances in technology finally made it possible to observe in spectral windows other than the visible range both from the ground and from space. The exploration through these new windows naturally turned into sky surveys, resulting in catalogs of radio, x-ray, and infrared sources. The fraction of extragalactic sources in these surveys have ranged from the substantial to the dominant.

All sky surveys have been instrumental in advancing astronomy. Catalogs are an invaluable tool to statistical astronomy, and a permanent source of surprises, of new classes of objects hiding among "normal" sources. Much is learned from the comparison of a new catalog to existing ones. For example, the comparison of IRAS Point Sources to optical catalogs of galaxies revealed ultra-luminous galaxies (Soifer *et al.* 1984); their comparison to radio surveys led to the infrared-radio correlation (Helou *et al.* 1985). However, with the accumulation of surveys and overlap in catalog membership, it has become exorbitant to operate only in terms of catalogs, carrying $(N-1)$ potential connections for each entry in each of N catalogs. In addition, this mode of operation impedes the efficient use of the information to gain physical insights, making it harder for instance to weigh evidence on positional coincidence against evidence on emission properties.

In parallel, the astronomical technical journals have been witnessing their own explosive growth in the publication rate. Abt (1988) estimates that the number of articles related to galaxies stood at 1,500 for the publication year 1985, and that it is doubling every eight years. Galaxies and cosmology are the fastest growing among eleven sectors of the astronomical literature surveyed by Abt.

The rend of increasing quantity and diversity of data will continue, driven by constantly improving astronomical instrumentation and data processing capabilities, and by large space missions such as the Hubble Space Telescope (HST), the infrared Space Observatory (ISO), the Advanced X-ray Astronomy Facility (AXAF), and the Space Infra-Red Telescope Facility (SIRTF), as well as ground-based large surveys (2MASS, Sloan, etc.). This trend makes it harder for individual researchers to stay up to date, in terms of being aware of new data, and tracking new ideas. This is the dual challenge posed by explosive data growth: dealing with the sheer volume, but also interconnecting intelligently the huge variety of information available.

in early responses to this challenge, attempts were made to consolidate the published results into reference catalogs (*e.g.*, the effort by de Vaucouleurs & de Vaucouleurs (1964) which culminated in the publication of RC3 in 1991), or more specialized compilations of data such as the Palumbo *et al.* (1983) Catalogue of Radial Velocities, the Huchtmeier and Richter (1989) General Catalog of HI observations, or the Catalog of Infrared Observations (Gezari *et al.* 1993). In the early 1980's the Centre de Données Stellaires of the Observatoire de Strasbourg introduced a new approach to dealing with data, namely a computer database service called SIMBAD (Set of Identifications, Measurements and Bibliography for Astronomical Data), which focused then on stellar data. The innovation was in the exclusively electronic archiving and interrogation, with continuous updating of the contents. Since then, a variety of systems and services have been set up, making up collectively a significant element of the astronomy research environment, as this book amply demonstrates.

in 1987, a fortuitous confluence of interests and opportunities, and the inspiring success of SIMBAD, prompted a number of IPAC astronomers to propose to NASA to establish a new extragalactic database that would be kept up to date in bibliography and published data, and be open to the astronomical community. IPAC as an environment was well suited to such an enterprise in terms of infrastructure, expertise, and experience with catalogs. Work on the design and implementation of NED started in June 1988, and its public release was announced at the 176th Meeting of the American Astronomical Society at Albuquerque in June 1990. NED has been in operation continuously since then.

10–3 Scientific Contents

NED is an object-oriented database, meaning that it organizes all information around individual extragalactic objects as opposed to leaving this information stored in catalogs or compilations; catalog membership is but one of many object attributes. In database management terminology, NED is organized as a relational database. It is built around a set of tables called the

Object Directory (3.1); this is the cornerstone of the database, to which additional tables are connected, usually via object identifications. These tables contain bibliographic references (3.3), published data of arbitrary description (3.4), notes from various sources (3.5), or copies of original catalogs (3.6). Such an architecture' is easily extensible, so that other types of information (such as images or spectra) can be added in the future.

3.1 OBJECT DICTIONARY

The Object Directory is the master list of objects recognized by NED, along with their positions, redshifts, and basic attributes (3.2). It represents the systematic merger of some 40 major astronomical catalogs, as well as scores of shorter lists culled from the refereed literature, such as the Palomar-Green list of quasars (Schmidt and Green 1983), the Schombert and Bothun (1988) low surface brightness galaxies, or the Hickson (1989) compact groups. New lists and catalogs are continually folded in based on thorough cross-identifications for each object made by NED team members. As of this writing (November 1994), the Object Directory contains about 330,000 objects known by 650,000 names. Naturally, NGC and IC names and a few hundred special names (e.g., Einstein Cross, Taffy Galaxy, etc.) have also been added to the Directory.

Table 10–1 shows some 20 catalogs with more than 4,500 members each, all of which have already been folded into the Object Directory. The abbreviations in the second column of Table 10--1 are standard nomenclature in NED, and are used in what follows to refer to these catalogs, Column 3 reports the number of objects from that catalog that appear in NED. This is often different from the count of entries in the catalog for two reasons: some of the catalogs include Galactic objects which are not folded into NED; and many catalogs include multiple objects appearing as single entries subsequently separated out by NED. The next set of challenges will be to fold into NED the large compilations of galaxies identified from automated scans of digitized optical sky surveys.

Most modern catalogs consist either of mostly Galactic or of mostly extragalactic sources (such as the PKSCAT90 radio catalog), or they will come with a relatively efficient prescription for distinguishing between the two populations (like the ESO/Uppsala or the *IRAS* Point Source Catalog). To appear in NED, an object has to have been classified as extragalactic in the literature, or to have a reasonable expectation (50%) of being extragalactic based on its observed properties. For instance, IRAS sources were selected for inclusion in NED only if (i) they had flux densities flat or rising with increasing wavelength, and (ii) were located in areas of the sky with low cirrus confusion noise (IRAS Explanatory Supplement 1987). Given such prescriptions, NED will unavoidably contain objects belonging to the Milky

Catalog Name	Abbreviation	Number of Entries
5 GHz Green Bank Survey	87GB	55,000
Galaxies in the Lick Proper Motion Catalog	NPM1G	50,517
IRAS Faint Source Catalog	IRAS F	49,011
1.4 GHz Green Bank Survey	[WB92]	30,239
Catalog of Galaxies and of Clusters of Galaxies	C(X'G	29,418
Morphological Catalog of Galaxies	MCG	29,003
ESO/Uppsala Catalogue of Galaxies	ESO	18,438
ESO Surface Photometry Catalogue	ESO-LV	15,467
"1'bird Bologna Radio Catalog	B3	13,353
Uppsala General Catalog of Galaxies	UGC	12,921
IRAS Point Source Catalog	IRAS	10,548
New General Catalog	NGC	7,021
Coma Cluster Catalog (Goodwin *et al.* 19&3)	[GMP83]	6,726
Galaxies Behind the Milky Way (Saito *et* al. 1991)	CGMW	6/509
Arp-Madore Catalog of Peculiar Galaxies	AM	6,445
Kiso Ultraviolet Galaxies	KUG	6,348
Parkes Catalog 1990	PKS	5,991
Southern Galaxy Catalog	SGC	5,481
Abell Clusters of Galaxies	ABELL	5,250
Shapley 8 Cluster Catalogue (Metcalfe *et al.* 1994)	[MGP94]	4,904
Fourth Cambridge Radio Catalogue	4C	4,843
Flat Galaxy Catalog (Karachentsev 1993)	FGC	4,754
Hewitt & Burbidge Optical Catalog of Quasars	[HB89]	4,617

Table 10-1: Major Extragalactic Catalogs in the Object Directory

Way. We find their inclusion preferable to the exclusion of extragalactic sources; objects re-classified after having been mistakenly called extragalactic will remain in NED to signal that switch. On the other hand, we will also miss some extragalactic sources, a price to pay to avoid including many more Galactic objects; this omission is remedied whenever detailed work is published reporting the extragalactic nature of such objects.

For each object, the Directory contains positional data, names and "basic data" (3.2), and a "preferred object type", *i.e.* the most useful description of this object by one of the categories in Table 10-2 (*e.g.* "galaxy" or "quasar" is preferred to "infrared source" or "ultraviolet excess source").

Sub-galactic objects (HII regions, globular clusters, etc) within other galaxies are now infrequently found in N} ID, but may be included eventually in a systematic fashion. For these, we use object type abbreviations based cm, and slightly simplified from, those used by SIMBAD (*e.g.* "SN" for supernova, "PN" for planetary nebula, de.). The same object type abbreviations are used for objects within the Milky Way, except that they are prefixed by an exclamation mark to emphasize their Galactic location.

Much care goes into the collection of positions and redshifts into NED, and they are continually over-written by more accurate values as they become available. Positions are stored internally in the J2000 system, along with their uncertainty, and a reference to their origin. All published positional uncertainties are transformed to a representation as a 95% confidence ellipse, whose semi-major and semi-minor axes anti position angle are kept. We also store Galactic coordinates, and total Galactic extinction in the blue at the position, derived from the Burstein-Heiles (1978) reddening r-nips. Ecliptic and super-galactic coordinates, and Equatorial coordinates other than J2000, are computed at display time.

3.2 NAMES AND BASIC DATA

The various names by which each object is known are stored and displayed in uniform NED formats, *e.g.* 4C +00.30 or UGC 00299, though users ma y use variations as a starting point for name searches. Each name is associated with an object type (see Table 10–2) which reflects the kind of survey that originated that name. Thus, even users not familiar with the specific surveys can see at a glance that UGC 12699 was recognized as a galaxy in eight different catalogs, and detected as an ultraviolet excess source (called MRK 0538), as an emission line source (UM 167), as an infrared source (1 RAS 23336+ 0152), and as a radio source (87GB 233340.0+ 015229). The discovery methods associated with most entries in Table 10–2 are self-evident; an "absorption line source" is one revealed by absorption against a bright continuum source, typically a quasar; a "visible-light source" refers to an image with insufficient

GClstr	Cluster of Galaxies
GGroup	Group of Galaxies
GTrpl	Triplet of Galaxies
GPair	Pair of Galaxies
G	Galaxy
QSO	Quasi-Stellar Object
RadioS	Radio Source
IrS	Infrared Source
VisS	Visible-Light Source
EmLS	Emission Line Source
AbLS	Absorption Line Source
UvES	Ultraviolet Excess Source
XrayS	X-ray Source
GammaS	Gamma-Ray Source
Other	Unusual object type
Q_Lens	Gravitationally Lensed Quasar
PofG	Part of Galaxy

Table 10–2: Extragalactic Object Types in NED

morphology for a definite classification as galaxy or star. “Other” sources are real objects so unusual or rare that they do not warrant a separate type definition, such as isolated intergalactic clouds detected in emission.

NED uses for object names acronyms as suggested by the authors, or as in current usage. However, when an acronym is not provided for new objects, NED finds it necessary to create one to provide unique identification. In general, the acronym then used consists of the first initials of the first three authors’ last names followed by the last two digits of the year of publication, with the whole string enclosed in square brackets; this is then followed by each object’s identifier as given in the paper, or each object’s coordinates as in the paper. Examples of such constructions are found in Table 10–1. NED also introduces suffixes to names of objects as required to resolve conflicting or overlapping identifications. For instance, a member of a pair identified only in the notes of the UGC would be referred to as UGC 08333 NOTES01; more commonly used are suffixes “NEDxx” when an ambiguity is resolved by NED, or “ID” when a radio or an infrared source is identified as a galaxy but no new name is provided.

“Basic Data” are the attributes most essential to a broad description of the object at hand. There is a different set of attributes for each object type, but each object has only one set of basic data, that corresponding to its “preferred object type". UGC 12699 for instance will only have basic data appropriate for a galaxy, namely an optical magnitude, a major and minor diameter, and a morphological description; On the other hand, a radio source has a flux density, the radio frequency of that measurement, a spectral index in the vicinity of that frequency, a size, and a morphology (*e.g.* head-tail); a quasar is described by an optical magnitude, and a qualifier (*e.g.* BL Lac m radio quiet).

These basic data are indicative values only, for they originate from many different sources not explicitly identified in the database. No attempt has been made to place them on a uniform scale. The main sources are catalogs and compilations, with the more accurate data sets favored, and the larger ones preferred at comparable accuracy. More rigorously defined and referenced data go into the photometry and positional data tables described in section 3.6.

Finally, "essential notes" generated by NED are attached to some objects (12,000 as of this writing) to point out significant facts, such as an erroneous identification, unique property, discordant value, or special relation to another object; these notes are always displayed along wit h basic data.

3.3 Bibliographic References

This segment of NED consists of pointers indicating the existence of useful information on a certain object in a given publication. A full bibliographic reference is kept for each publication in addition to a 19-character reference code which encapsulates the full reference in abbreviated form (see chapter 24 in this book).

The bibliographic references in NED derive from two main sources. Starting in 1988, members of the NED team have been reading systematically several of the major journals to identify papers presenting meaningful new information on extragalactic objects, as we] 1 as papers of extragalactic interest in general. This coverage is highly reliable and complete for *A&A, A&AL, A&AS, AJ, ApJ, ApJL, ApJS, MNRAS,* and *RASP* starting in 1990, for *IA U Circulars* starting in 1991, and for *Nature, PASJ, Asf ronomy Reports* and *Asf ronomy Letters* (the last two formerly *Soviet Astronomy* and its *Letters*) starting in 1992, This segment of the literature is now yielding about 75,000 pointers per year. The second source of bibliographic data is the SIMBAD project, which has kindly provided all of their references to extragalactic objects up to 1989, and has been providing updates on an annual basis since then. The SIMBAD pointers are complementary, since they are based on a search of many more astronomical journals conducted for many years by the librarians at the Institut d'Astrophysique de Paris, and which has produced, for extragalactic sources, systematic coverage starting in 1983, and sporadic coverage going back to 1917.

As of this writing, the NED database contains well over 500,000 pointers linking objects to over 25,000 distinct publications. [bibliographic pointers to the journals scanned by NED are typically available on-line about one month after the corresponding issues have appeared in print.

3.4 PUBLISHED MEASUREMENT DATA

One of the goals of NED has always been to collect and store information about new extragalactic data appearing in the literature. The goal was to carry fully those data that can be expressed in a few numbers, and have clear descriptions of those which cannot, such as spectra and maps. This ambitious long-term goal is being undertaken in steps, with the treatment of photometric data being the first major implementation to date.

Broad-band flux densities, fluxes and magnitudes at any wavelength, and fluxes in the 21 cm HI line and the 2.3 mm CO line are now routinely entered into the database, mostly from large compilations and catalogs, working towards an eventual systematic coverage of the literature. The effort concentrates exclusively on measurements of global or nearly global emission from objects. Each measurement is fully referenced and cast internally into a uniform "data frame" which includes the most significant information needed for a critical appraisal, *e.g.*:

- the as-published wavelength or frequency, flux value, units and uncertainty, m upper limit, and the meaning of the last two quantities, such as "2σ" or "plate limit";
- indications pertaining to the derivation in the spatial domain, such as "flux in fixed aperture", or "integrated from a map"; similarly for the frequency domain, such as "synthetic band" or "integrated over line";
- a variety of frequently recurring indications of a general nature, such as "from new raw data", "homogenized from previously published data", "extinction-corrected for Milky Way", or "K-correction applied".

NED currently offers about 635,000 photometric data frames covering the spectrum from the radio to the X-rays. As each measurement is integrated into NED, its equivalent value on a uniform system of units is also stored, making it possible to compare all data across objects and across the spectrum. Building on these coherent data, NED will soon offer its users a view of the broad-band spectral energy distribution using available data across the whole observable spectrum for any object.

Besides photometry, current plans call for similar data frames to be collected for position and redshift measurements, morphological classification and spectra] or other typing.

3.5 NOTES

Almost every catalog published has a wealth of data appearing as notes on individual objects, usually in an appendix that does not get circulated in computer-legible form, and therefore remains largely untapped. Many journal articles also contribute valuable notes which are all too easily overlooked or forgotten. NED has made a special effort to make notes available. In some cases, this has entailed digitizing the printed material for the first time (*e.g.* the Hubble Atlas of Galaxies (Sandage 1961), all 8,000 UGC notes (Nilson 1974)). In other cases, *e.g.* the ESO/Uppsala and the SGC, the shorthand used in the notes was translated into regular English before the notes were included in NED; in the case of the MCG or the Arakelian list (1975), the contents of the notes were translated from Russian to English before entry. As of this writing, the database contains over 34,000 individual notes. The notes are retrievable by query on the object name, and are stored along with a reference to their source, and the particular object name used by the author of the note.

3.6 LITERAL CATALOGS

The major catalogs (such as those listed in Table 10-1 above) are the source material from which NED assembles the Object Directory, and as such they need to be accessible in their original form, especially if they contain substantially more information than NED has extracted from them. Under the title of “literal catalogs”, NED provides users with a view of the catalog entries as close as possible to their original printed appearance. Currently available are images of the RC3, UGC, ESO/Uppsala, PKS, and [HB89].

3.7 ABSTRACTS OF PAPERS AND THESES

In the course of scanning the core journals (3.3), papers of extragalactic interest have their abstracts digitized by the NED team and integrated into NED as text, and made immediately available to users for browsing. Besides articles that contribute original data, the abstract collection includes reports of theoretical studies, modeling, or empirical analysis. NED currently offers about 10,000 abstracts on-line.

NED introduced in 1992 an on-line collection of dissertation abstracts of extragalactic interest. Thesis abstracts (mostly from U.S. institutions) almost complete back to 1980, and titles and authors dating back to 1909 were generously provided by University Microfilms. New thesis abstracts are immediately added to this collection once they have been accepted by the granting institution and forwarded to NED.

10–4 **Functions**

The user's view of a database is defined by the data retrieval process. From this point of view, NED can be primarily described as a source of lists of objects and lists of bibliographic references, with several ancillary functions attached. There arc many ways to construct lists of objects, which divide into two modes, local searches and global searches (4. 1). Having retrieved a list of objects, members in that list can be singled out for follow-up, typically retrieving lists of relevant references (4.2), or of various types of data pertaining to these objects (4.3). Lists of references can also be constructed directly. It is also possible to branch in cascade from lists of objects to lists of references, although the interface will only have available the most recent list of each kind. Ancillary functions provided by NED include the plotting of the distribution on the sky of objects in lists, the browsing of paper and thesis abstracts, a coordinate conversion and precession utility, various information files (e.g. a list of upcoming conferences), session history tracking, and the transmission to the user by electronic mail of ASCII files containing data retrieved during a session.

Aside from these functions, there is an elaborate collection of software tools, largely invisible to the user, whose purpose is to populate NED with new objects and data and update existing information, while preserving data integrity and consistency, and maintaining traceability of the modifications.

4.1 SEARCHING FOR OBJECTS

Under the local search mode, users can locate objects by specifying a name, or a vicinity to search in. Substantial latitude is allowed in entering namesname resolver, with the input being interpreted by the NED interface and cast into the standardized formats used by NED for internal storage (section 5). Moreover, a search by name (*e.g.*, VV 136) will normally return (unless otherwise selected by the user) all objects that carry the specified name plus a suffix (*i.e.*, VV 136, VV 136a, VV 136b, and VV 136c). Search by vicinity returns all objects lying within a circle, whose radius and center are entered by the user. Such searches can be centered on an object specified by name, in which case the occurrences of suffixed versions of this name are ignored. If the center is specified by position, it can be given in any of four coordinate systems: Equatorial or Ecliptic (arbitrary equinox and epoch years), Galactic, or Super-Galactic. Searches by vicinity can include additional filters on object parameters, such as including or excluding object types, and restricting redshifts to certain intervals. in another local search mode, the user enters an abbreviat ion of coordinates in the "IAU style" commonly used to construct names from positions (e.g. 2254-054), and a positional search is made by NED covering the most likely area in the sky from which these coordinates might have originated.

The simplest global mode search retrieves all objects linked to a given reference. The most recent and powerful of the search modes allows users to search through the whole database by parameter, e.g. to find "all quasars at Galactic latitudes between 45" and 50° and with redshifts greater than one", or "all clusters of galaxies which are not known to be X-ray sources". Searches can now be made for objects satisfying simultaneously conditions on Equatorial and Galactic coordinates, redshift and object types; constraints on photometric data and other attributes wi 11 be added incrementall y. These conditions are specified in terms of (i) coordinate intervals to include m exclude, (ii) of object types to include or exclude, and (iii) for redshifts (since not all objects have them) in terms of availability or lack thereof, as well as included or excluded intervals.

Each one of those searches returns a list of objects, along with the corresponding contents of the Object Directory and Basic Data (3.1 and 3.2 above), and with the number of bibliographic references (3.3), the number of photometry data points (3.4), and the number of notes (3.5) available for each object, thus presenting the user with a summary status of the object within NED. The interface displays this list, allowing the user to examine in detail individual objects and retrieve further information as detailed in the next section.

4.2 RETRIEVING DATA

Data retrieval requests presently in operation correspond to those data structures described in section 3 above. Users can list the bibliographic references for specific objects; once a reference has been found, its abstract can be displayed. Users can browse through the notes and the literal catalog entries linked to individual objects. Similarly, photometric data frames can be listed for each object, and examined in detail.

Requests for notes and references can be issued to follow up on an object identified in a previous search for objects, or can be formulated *ab initio* by specifying an object name. The second mode offers (as the default option) the advantage of an "extended" search, which starts by locating other objects which have a name root in common with the name specified, then retrieves all references to all objects thus related to the object of interest. For instance, the Hickson Compact Group HCG 23 consists of five galaxies called HCG 23A through HCG 23E. NED contains notes pertaining to, and references linked to, only some of the five individual members. A search for articles strictly relevant to HCG 23 will not reveal papers linked to individual member galaxies, whereas an extended search will return all references to HCG 23 as we] as HCG 23A through HCG 23E. More useful yet, the same all-inclusive list will be returned b y an extended search on NGC 1 ? 16 which also happens to be the galaxy HCG 23C.

10–5 Interfaces

The functions described above are provided via three distinct interfaces, all of which are served by the same underlying data and software. The direct human-machine interface (5.1) offers the most versatile and detailed mode of extracting information, and typically drives the design of new capabilities. However, most functions are also available in simpler forms in a batch mode via e-mail (5.2), and in a client-server mode for machine-to-machine interaction (5.3). There is no charge for logging into the service or retrieving data from it in any of the modes.

5.1 HUMAN-MACHINE INTERFACE

The user's view of a database is strongly colored by the ease with which one interacts with it. This case is defined by several factors, including: (i) the amount of learning needed before sessions become productive; (ii) the degree of overlap between the specific questions a user has and those the database can answer; (iii) the power and flexibility available to the user in formulating a query; (iv) the convenience in the mechanics of submitting a search and collecting the results; and (v) the elapsed time between submitting a query and gettting the response. While some of these factors are determined by the internal structure of the database, they can be modified by, and in the end depend mostly on, the user interface to the database.

in line with those perceptions, NED's primary interface was designed as an interpreter that enables an astronomer or a librarian to use the database without learning its jargon, understanding its internal data organization, or being familiar with the interface mechanics. One of its main goals is to make a user's first session productive in terms of obtaining the sought-after data within a reasonable amount of time. This goal has been dubbed the "five minute rule": unless the interface yields useful data within five minutes of a new user's first session, that user might never return to the database.

The direct NED interface may be accessed by setting up a network connection to IPAC from any Internet host, using a command of the form <telnet ned . ipac . cal tech. edu>. Once connected to the NED platform and prompted for a "login", the user should respond with "NED"; no password is needed. The interface then gives the user a choice between a character-based VT100 terminal mode or an X-Window-based graphical interface with plotting capabilities. These two modes of presentation run the same source code, compiled with different options, and access the same data. The user can then choose between them based on the hardware or software available at their end, and on the bandwidth of their connection to IPAC. The X-Window mode offers advantages in clarity of presentation and plotting abilities (sky distribution, spectral energy distribution), whereas the character-based mode requires a fraction of the network bandwidth to run at an acceptable pace.

Users control the NED session by selecting options and entering data within the context of a menu tree; the interface presents itself as a screenful Of text for each node in this tree, with a standard screen format maintained throughout the session. This format consists of three areas: the one at the top displays available options, while the one at the bottom displays commands, and the middle box serves for input and output. Options are displayed as abbreviated descriptions of the functions they serve. As an option is selected (typically with a single key stroke or a cursor point-and-click), a new screen is displayed, a data input screen is activated, or a data output sequence is initiated. The commands perform simple, general purpose tasks such as terminating input or output, moving to a higher-level menu, obtaining detailed help, or ending the session.

A user would start a session, then select options until the desired input panel is reached; once the various fields in the selected panel have been filled out, a corresponding search is submitted, and the interface is inactive until the search concludes. For potentially time-consuming searches, the user may be warned and asked for verification before the search is submitted. The resulting data are then available for display in various output screens, until a new search is requested, causing the new results to overwrite the previous ones. Users can request all data returned during a session (including abstracts) to be sent to them by electronic mail.

An essential feature of the NED interface is self-documentation. In the first place, a NED session is steered at every step by a choice among options expressed in astronomical jargon, instead of a database interrogation language. Secondly, help functions are available at several levels throughout a session: (1) All menu items, options and input fields are labelled in simple terms, and every error message informs the user directly about what could not be processed. (2) A longer explanation of the functions and usages within every screen can be displayed by selecting the "HELP" option within that screen. (3) An introduction and overview, a general tutorial and news about recent additions to the system are also available as menu options in the first NED screen.

Similarly extensive documentation is available on output; data fields are labeled in whatever detail is necessary, and abbreviations are avoided as much as possible. On the screen displaying names of objects (cross-identifications) for instance, the user can ask for information on a given name, and get a brief description and a bibliographic reference to the catalog in which the name originates. In addition, users can tailor their output environment by selecting many of the display attributes, such as parameters for sorting lists, or coordinate system for displaying positions.

As the NED interface is accepting data from a user, it checks them to the extent possible for ambiguities, conflicts or errors; more significantly, it attempts to interpret them, and echoes its interpretation to the user for verification. These verification capabilities translate into less restrictive formats for the user to adhere to; for instance, the right ascension field will accept any of "6h0m5s", "6:0:5", or "6 O 5" as valid input, and interpret it as "06h00m05s".

The most versatile and sophisticated interpreter in NED is associated with names of objects, since naming conventions show great diversity across the literature. This interpreter is based on the "lex" regular expression processor; it gets the nomenclature guidelines out of a file, thus simplifying the task of updating name conventions recognized by NED. It usually generates one of four responses: the input is uniquely recognized, and rewritten in the standard NED internal format in preparation for querying the database; the input is ambiguous, so a number of possible interpretations are offered for the user to choose among; the input is distantly related to one or more naming conventions, so the proper formats for the latter are offered to the user as information; or the input is completely unrecognized by NED. Examples of each case are: "m1 -2-3" is accepted as "MCG + 01 -02-03"; "sculptor" returns a choice between "Sculptor Group", "Sculptor Galaxy", "Sculptor Dwarf Elliptical", or "Sculptor Dwarf Irregular"; and "qso 1234" will suggest nomenclatures of the form " [HB89] HHMM+DDda" or "QS FN:NN" with explanations of their origin.

To underline the importance of the users' suggestions, every screen offers a menu option for users to leave comments on the system. Comments are reviewed and acted upon almost daily by the NED team. They include possible errors in the data, requests for new functions or data, and quality ratings of the service.

5.2 BATCH PROCESSING MODE

In this mode, a user e-mails a set of queries to `nedbatch@ipac.caltech.edu`, to be processed in a queue without human intervention. The results are written to file on a public access directory, and the user is notified by e-mail of the status of the processing, and the location and size of the results file. The queries have to be submitted following a specific format, a model of which will be mailed upon request by the primary interface.

5.3 SERVER MODE

This is the most recent interface to be offered by NED, and has therefore a more reduced access to NED data. It is designed to allow software developers to integrate a NED query into a computer application in a manner that requires no intervention from the user. For instance, a database of images accessible only by position could be enhanced by allowing users to enter names of galaxies, then have the software query NED directly for the position attached to the name, and use the resulting information to address the database; this provides a major enhancement at very low cost in development.

10–6 Architecture

The NED software resides on two platforms, a SUN Sparcstation 10 which supports all the user interface functions and the abstract database, and a SUN4 server which carries the search functions, and where most of the data are kept and managed. On the interface platform, the main software module which manages the session was written expressly for NED, but relies for screen input and output on a commercial package called JAM (for "JYACC Application Manager "), developed and marketed by JYACC. On the server side, all database management system functions are handled by SYBASE, a commercial software package that largely follows the relational database paradigm. Communication between platforms is handled by SYBASE client-server connections over the local IPAC network. This architecture offers many advantages in terms of performance, security, and manageability, and makes optimal use of the available platforms.

10–7 The Future

NED is still growing, both with new functions being added, and old ones becoming more sophisticated. in addition, the contents of NED continue to grow with more catalogs being folded into the Object Directory, more pointers to the literature being entered, more data — photometric and otherwise — being added, and more abstracts becoming available. in the long term, NED will certainly undergo much evolution, for it was designed with flexibility and ease of upgrades in mind. Enhancements presently envisaged call for more visualization (plotting and text presentation) capabilities on the user interface, and an expanded suite of functions accessible via the server mode. Exploratory efforts are also under way in collaboration with J. Mazzarella to establish a Mosaic-based interface for NED to take advantage of the capabilities offered by the World-Wide Web system.

This is a time of transition, with the information revolution finally changing the way astronomers conduct their business, and electronic publishing becoming a reality. By their nature, database services are both symptoms and catalysts of that transition. NED is already playing its role in this arena, although it relies on the existing peer-review system for ensuring data quality, drawing its data primarily from the refereed literature. However, NED goes beyond these foundations by offering new modes of tapping into the literature; the growing collection of interconnected data, references and abstracts will make possible yet more innovative modes. By greatly simplifying certain queries, NED leads to new habits in research; projects that used to be dauntingly labor-intensive can now be undertaken by a lone researcher spending an afternoon at a terminal. Exploratory work that would have seemed "too risky" because of its considerable cost suddenly is worth the effort. How does NED contribute to astronomical research? By making it affordable to ask questions, NED paves a new high road to discovery.

Acknowledgments

Many people have contributed to NED, and continue to do so; we would like to thank Rick Ebert and Maneesh Sahani for their help in the software arena; we also recognize IPAC's support, especially through the computer systems group. We especially acknowledge the friendly collaboration of the SIMBAD project, Centre des Données astronomiques de Strasbourg. Last but not least, we thank all the authors and the National Space Science Data Center, who send us computer-readable versions of their data or catalogs.

NED is funded by the Science Operations Branch of the Astrophysics Division, Office of Space Science and Applications, NASA. This work is carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

[1] Abell, G.O., Corwin, H.G., Jr., and Olowin, R.P. (1989), ApJS, 70,1. (ABELL)

[2] Abt, H.A. (1988), PASP, 100,1567.

[3] Arakelian, M.A. (1975), Soobshch. Byurakan Obs. Akad. Nauk. Arm. SSR, 47, 1. (ARK)

[4] Arp, H. C., and Madore, B.F. 1987, "A Catalogue of Southern Peculiar Galaxies and Associations" (Cambridge: Cambridge University Press) (AM)

[5] Becker, R.H., White, R. L., and Edwards, A.L. (1991), ApJS 75,1. ([BWE91])

[6] Burstein, D. and Heiles, C. (1978), ApJ, 225,40.

[7] Caswell, J. L., and Crowther, J.H. (1969), MNRAS, 145,181. (4C)

[8] Corwin, 11. G. Jr., de Vaucouleurs, A., de Vaucouleurs, G. (1985), "Southern Galaxy Catalogue". (Austin, TX: University of Texas) (SGC)

[9] Dreyer, J.] ..E. (1888), MemRAS 49,1. (NGC)

[10] Gezari, D. Y., Scemitz, M., Pitts, P. S., and Mead, J.M. (1993), "Catalog of Infrared Observations", NASA RI'- 1294. (CIO)

[11] Goodwin, J. G., Metcalfe, N., and Peach, J.V. (1983), MNRAS, 202, 113. ([GMP83])

[12] Gower, J.F.R., Scott, 1'.1'., and Wills, D. (1967), MmRAS, 71,49. (4C)

[13] Gregory, I? C., and Condon, J.J. (1991), ApJS, 75,1011. ([87GB])

[14] 1 lelou, G., Soifer, B.T., and Rowan-Robinson, M. (1985), ApJL, 298, L7.

[15] 1 Iewitt, A., Burbidge, G., (1989), ApJS, 69,1 ([HB89])

[16] 1 lickson, P, Kindl, E., and Auman, J.R. (1989), ApJS, 70, 687.(} ICG)

[17] I Iuchtmeier, W. K., and Richter, O.-G. (1 989), "A General Catalogue of 1 II Observations of Galaxies" (New York: Springer-Verlag)

[18] IRAS Point Source Catalog, Version 2 (1 988), Joint 1 RAS Science Working Group (Washington, DC: US GPO) (IRAS)

[19] Karachentsev, I. (1 993), AN 314,97. (FGC)

[20] Klemola, A., Hanson, R., and Jones, B. (1987), AJ 94,501. (NPM1G)

[21] Lauberts, A. (1982), "The ESO/Uppsala Survey of the ESO(B) Atlas" (Garching bei Munchen: European Southern Observatory) (ESO)

[22] Lauberts, A. and Valentijn, E.A. (1 989), "'['he Surface Photometry Catalogue of the ESO-Uppsala Galaxies" (Garching bei Munchen: European Southern Observatory) (ESO-LV)

[23] Metcalfe, N., Goodwin, J. G., and Peach, J.V. (1994), MNRAS, 267, 431. ([MGP94])

[24] Moshir *et* al. (1992), "Explanatory Supplement to the IRAS Faint Source Survey, Version 2", JPL IJ-1OO15 (Pasadena: J PL) (IRAS F)

[25] Nilson, P.(1973), "Uppsala Genera] Catalogue of Galaxies", Acts Universitatis Upsaliensis, Ser. V:A vol. 1 (Uppsala: Reg. Soc. Scient. Ups.) (UGC)

[26] Nilson, P. (1974), "Catalogue of Selected non-UGC Galaxies", Uppsala Astronomical Observatory Report No.5. (UGCA)

[27] Palumbo, G. G.C., Tanzella-Nitti, G., and Vettolani, G. (1983), "Catalogue of Radial Velocities of Galaxies" (New York: Gordon& Breach)

[28] ParkesCatalogue(1990), the Southern Radio Source Database, ver. 1.01, compiled by A. Ii. Wright and R.E. Otrupcek. Magnetic tape and preprint, Australia Telescope National Facility. (l'KS)

[29] Pilkington, J. D.H., and Scott, P.F. (1965), MmRAS, 69,183. (4C)

[30] Saito, M. *et al.* (1991), PASJ, 43,449. (CGMW)

[31] Sandage, A. (1961), "The Hubble Atlas of Galaxies" (Washington, DC: Carnegie institution of Washington)

[32] Schmidt, M., and Green, R.F. (1983), ApJ, 269,352, (PG)

[33] Schombert, J. M., and Bothun, G.D. (1988), AJ, 95,1389. ([SB88])

[34] Soifer, B.T. *et al.* (1984), ApJL, 283,1.1.

[35] Takase, B., and Miyauchi-Isobe, N. (1993), Pub]. Nat]. Obs. Japan, 3, 169. (KUG)

[36] de Vaucouleurs, G., de Vaucouleurs, A., and Corwin, I-I. G., Jr. (1976), "Second

Reference Catalogue of Bright Galaxies" (Austin: University of Texas Press) (RC2)

[37] Vorontsov-Velyaminov, B. et al. (1962 74), "Morphological Catalogue of Galaxies", in 5 parts (Moscow: Moscow State University) (MCG)

[38] White, 1<.1.., and Becker, R.I 1. (1992), ApJS, 79,331. ([WB92])

[39] Zwicky, F. et al. (1961-68), "Catalogue of Galaxies and of Clusters of Galaxies", in 6 volumes (Pasadena, CA: California Institute of Technology) (CGCG)

Access Pointers

[-1] NED WWW page:
http: //www.ipac.caltech.edu/ipac/projects/ned.html

[-2] NED telnet access: telnet://ned@ned.ipac.caltech.edu/