PREPARING A PUBLIC DATABASE OF RADIO SOURCES

H. Andernach\(^1\), S. A. Trushkin\(^2\), A. G. Gubanov\(^3\), O. V. Verkhodanov\(^2\), V. B. Titov\(^3\) and A. Micol\(^4\)

\(^1\)Departamento de Astronomía, University Guanajuato, Apdo. Postal 144, Guanajuato, C.P. 36000, Mexico
\(^2\)Special Astrophysical Observatory, Nizhnij Arkhyz, Karachay-Cherkess Republic 357147, Russian Federation
\(^3\)Astronomical Institute, St. Petersburg State University, Peterhof 198904, Russia
\(^4\)ST-ECF, Karl-Schwarzschild-Str. 2, D-85748 Garching, Germany

Received July 20, 1996.

Abstract. We have collected the largest existing set of radio source lists in machine-readable form: 320 tables with 1.75 million records. Only a minor fraction is accessible via public databases. We describe our plans to make this huge amount of heterogeneous data accessible in a homogeneous way via the World Wide Web, with reliable cross-identifications, and searchable by various observables.

Key words: databases – radio sources – catalogs

Surveys of radio sources have always been important in widening our horizon of the distant Universe: optical identifications of the sources are pointing us to the most distant galaxies known, and the perfect isotropy of the source distribution in even the deepest surveys shows that we are sampling a much larger volume of space than we do with magnitude-limited galaxy surveys. Of course, an efficient search for a certain types of objects (e.g. distant galaxies, variable sources, etc.) requires the filtering of large samples of radio sources, carefully selected from literature by very specific criteria. Until now this involved constructing these samples by tedious merging and sorting.
of the numerous catalogs of individual sources. In addition, existing astronomical databases show a notable lack of published information about radio sources, and data centers provide only a small part of the largest published source catalogs.

One of us (Andernach 1990) noted this deficiency years ago. He has now gathered the most complete set of radio source catalogs to date in electronic form: 320 source tables with a total of 1.75 million entries. Only a minor fraction of the tables is accessible through public databases. This collection is a result of six years of persistent requests to the authors, even though in an early email campaign radio astronomers were invited to contribute their electronic source tables to the astronomical data centers voluntarily. In an effort to recover in addition the largest published source lists not available in electronic form, H. A. employed a scanner and the “Optical Character Recognition” (OCR) software. Since early 1995, over 100 printed data tables with > 50,000 entries have been converted into electronic form. Staff at the Special Astrophysical Observatory of the Russian Academy of Sciences (SAO RAS) also helped in correcting, editing and proof-reading the raw OCR result, as well as in the manual retyping of many source tables which could not be read with OCR equipment. Tables published only on microfiche are the most difficult ones to convert, since microfiche readers have turned into an extremely rare species of apparatus.

Independently, SAO RAS and its St. Petersburg branch, together with the Astronomical Institute of the St. Petersburg State University, maintain a collection of astronomical catalogs. The database of radio astronomical catalogs, RAC, has been created at the Astronomical Institute (Gubanov & Titov 1996; http://www.aispbu.spb.su/WWW/RAC.html). The CATalogue Support System (CATS) is being developed at SAO (Verkhodanov & Trushkin 1995; Telnet: ratan.sao.ru, login as cats, pw=<CR>). Both are supported by the Russian Foundation for Basic Research.

A. Micol is involved in the use of relational databases for astronomical applications. He will be responsible for the design of the database structure and its user interface, the translation of user to software requirements, the configuration and maintenance of a WWW server and for the Application Program Interfaces (API) to commercial database management systems.

The present project joins the efforts of these groups in establishing the first reasonably complete and publicly accessible database of radio sources. During the first half year of active collaboration over
fifty tables and catalogs were either scanned or typed, proof-read and included in the full collection of data sets. Tools were developed to cross-identify sources from different catalogs and to construct and display their radio spectra.

Virtually all catalogs have a different format and list different observables. It will be a major challenge to provide uniform access to such a heterogeneous collection of data sets involving different methods, notations and units. Only a clear understanding of observing techniques and data-processing methods for a given catalog will guarantee reliable cross-identifications between catalogs. Thus, documentation files in an agreed format will be prepared for all catalogs accessible to the search. Many catalogs provide only source names and will have to be supplemented with positional data. For tables lacking positional errors we will insert these errors from formulae given in the publication.

The concept of a “Reference Directory” (a central repository of metadata like the descriptions of catalog fields, their physical units, mapping of original field names to the actual name in the database, etc.) will be used for catalog browsing and to process user queries.

The basic user requirements of the proposed database consist of
- being freely accessible via Telnet and World Wide Web;
- inclusion of a sufficiently complete number of source catalogs with on-line documentation of their columns;
- search by position in \((\alpha, \delta)^{1950}\), \((\alpha, \delta)^{2000}\) and Galactic coordinates;
- SQL-type query by user-specified parameters like e.g. name, flux at a given frequency, source size, spectral index, identification;
- X-window protocols for graphical tools like spectral plots, radio-sky finding charts and display of two-dimensional FITS files.

The proposed database will fill a long standing gap in the multi-waveband astronomical information systems at a time when several large-area surveys of high sensitivity and resolution are in progress with the VLA and the WSRT, which produce millions of further sources. Comparison of the available radio data with these new surveys will be crucial for statistical studies, and it will undoubtedly lead to the discovery of new source populations.

Cross-identification and spectral classification of the many 10,000s radio sources in our Galaxy and in extragalactic space will provide an unprecedented base for radioastronomical and cosmological studies. Scientific applications of such a database are extremely numerous, e.g. the search for high-redshift radio galaxies, the study
of the influence of a cluster environment on radio sources, the selec-
tion of targets for a future space VLBI mission with extremely high angular resolution and the selection of spectrally peculiar or
time-variable sources for follow-up studies. A much better separa-
tion of the complex mix of source populations in the Galactic plane will also be possible. The project will have its educational impact on students preparing catalog documentation and user’s guides, and collaborating in specific research projects.

REFERENCES

Andernach H. 1990, Bull. Inf. CDS, Strasbourg, No. 38, 69
Verkhodanov O. V., Trushkin S. A. 1995, SAO RAS Preprint No. 106, p. 66