

PREPRINT DATABASES AT NRAO AND STScI

Ellen Bouton
National Radio Astronomy Obs.
Edgemont Road
Charlottesville, VA 22903 USA

Sarah Stevens-Rayburn
Space Telescope Science Inst.
3700 San Martin Drive
Baltimore, MD 21218 USA

HISTORY AND DEVELOPMENT OF SYSTEM

A couple of different facts coalesced in the mid-70s that forced us to try to come up with a manageable scheme for handling preprints. First, as librarians at the National Radio Astronomy Observatory, we were charged with tracking papers by NRAO staff and by visitors using NRAO telescopes to provide an annual listing of staff and visitor publications and to aid in compiling statistics on telescope use.

Second, staff and visitors kept asking for preprints –by series and/or by author, subject, or cover color– and the number of preprint series being received had increased to the point where we could no longer *always* pull that information off the tops of our heads. So, after consulting with colleagues who had similar concerns, we put together a database in 1978, modestly based on the highly successful Preprint/Antipreprint lists produced by the Stanford Linear Accelerator Center. We decided to produce a biweekly list of new papers received and an as-needed list of recently published papers. The scientific staff were very enthusiastic from the beginning, seeing that our efforts would make their efforts to keep up with the literature that much easier.

The database was set up on an IBM mainframe with separate fields for (a) control number indicating year and biweekly period, (b) institutional abbreviation and preprint number, (c) authors, (d) title, and (e) citation.

The advent of the biweekly RAPsheet (Radio Astronomy Preprints) in mid-1978 was greeted with great enthusiasm by staff, as were the lists of previously announced preprints with their citations (the unRAPsheets) and the weekly printout of the entire preprint database by author/title used as a reference source in the library. We also produced regular lists by institution of preprints received so that preprints in series could be identified without having to keep them all after they had been published. Online searching of the database was extremely cumbersome, so we did virtually none in the years the database was on the IBM mainframe.

In mid-1983 when SS-R went to the Space Telescope Science Institute, she carried with her a tape of the database and spent four months setting it up on the VAX and getting the software written to produce the lists, known there as the STEPsheet (Space Telescope Preprints). The VAX allowed for online searching and other niceties that the IBM did not.

At the end of 1985, NRAO replaced the IBM mainframe with a Convex, keeping the existing VAXs. The RAPsheet went on the VAX using INMAGIC, database management software designed to handle library and bibliographic information. NRAO now had much more flexibility in the ways to use the database, added some fields (e.g., code for NRAO telescope used), and had online searching both in the command language version and using an in-house designed menu-driven system.

MECHANICS

Incoming preprints are added to the database as received, then displayed with incoming mail. Those that are part of a regular numbered series are also checked in as exchanges.

NRAO database record looks like this:

ID 8803001

INST/1 NRAO-88/13

INST/2 STSI-244

INST/3 JPL-242

AU/1 ANTONUCCI, R.R.J.; ULVESTAD, J.S.

TI/1 A Large New Family of Compact Radio Sources in the Starburst Nucleus of NGC 253

CITE Ap. J.

NRAO 4

PROP

PS u

STScI record looks like this:

88-05#STSI-244;JPL-151#ANTONUCCI, R.R.J.; ULVESTAD, J.S.#A large new family of compact radio sources in the starburst nucleus of NGC 253#Ap. J.\$

When the mail is put away, preprints are displayed in alphabetical order by first author in separate shelving section.

The tables of contents of *all* astronomy journals, general science journals (*Nature*, *Science*, etc.), and meeting proceedings are cross checked with a list of unpublished preprints to find citations. Other journals are scanned for astronomy or instrumentation papers. At NRAO, the author index of each *Physics Abstracts* issue is also run against the list of unpublished preprints; citations for approximately two papers per PA issue are picked up, usually from journals or meeting proceedings not received in the library or for preprints received after publication. Newly published papers are pulled from the display shelves and citations added to database; codes are changed to indicate the paper is newly published rather than unpublished. Although this is very labor-intensive, one needs to remember that most of this was being done anyway in tracking NRAO papers, without the benefits provided by having the database.

Records for published papers then appear like this:

ID 8803001

INST/1 NRAO-88/13

INST/2 STSI-244

INST/3 JPL-242

AU/1 ANTONUCCI, R.R.J.; ULVESTAD, J.S.

TI/1 A Large Family of Compact Radio Sources in the Starburst Nucleus of NGC 253

CITE Ap. J. 330: L97-L100, 1988

NRAO 4

PROP

PS n

88-05#STSI-244;JPL-151#ANTONUCCI, R.R.J.; ULVESTAD, J.S.#A large family of compact radio sources in the starburst nucleus of NGC 253#Ap. J. 330: L97-L100, 1988*

Biweekly RAP and unRAP sheets are produced and distributed to NRAO staff and to approximately 85 individuals and libraries outside NRAO. After distribution, code N for newly published papers is changed to P for published. Biweekly STEP sheets are copied and stuffed in mailboxes of those who want hard copy and distributed electronically to individuals at STScI as well as to places as far away as Hawaii, Argentina, and England. As at NRAO, the outside distribution has expanded primarily because staff who move on want to take the service with them. Biweekly STEP and RAP sheets are offset by one week. They are exchanged between the two libraries by e-mail, allowing each of us to edit the other's current list and append to our own a list of preprints received only at the other's institution.

ANNUAL HOUSEKEEPING

At the end of each year, published preprints received in the preceding year are off-loaded into a separate file; we have kept these files since they are useful for searching by key word, or while waiting for the Ap. J. annual index to appear. Thus, in 1988, the database contains all unpublished preprints, regardless of receipt date, and all published papers (with citations) received in 1988 and 1987. AT STScI, all STScI papers, published and unpublished, are kept in the main database; at NRAO this is not done, but one could pull out NRAO papers from the older database if needed. We both produce yearly listings by institution for the regular numbered preprint series, although we are not sure they are much used now that online searching is so easy. We also remove all unpublished preprints older than two years from our main display areas, although we do keep them in a separate area.

To give some idea of database size, the NRAO database, before 1986 papers were off-loaded at the end of 1987, had 3744 papers in it, of which 1454 were unpublished. On 20 July 1988 there were 3207 papers in the database, of which 1602 were unpublished. STScI figures are very close: the list in December 1987 had 4018 total and 1435 unpublished, and on July 19 had 3792 total and 1502 unpublished. Total differences are probably attributable to the 1987A database that is merged with the STEPsheet, since all 1987A papers are added whether a preprint is received or not.

COMMENTS

It takes a lot of time, but is one of the most heavily used parts of the entire library collection. Visitors from other institutions make a point of allowing time to browse through the collection. Staff remain wildly enthusiastic about the cumulated printout as a reference tool – they look for a published paper there before looking in the journal indexes. The online system is searched heavily, even by those who don't get the biweekly lists. In addition to helping the staff, it also helps the librarians enormously in keeping up with meetings and being able to tell people with a fair degree of certainty that if it appears as unpublished in the database, it is really likely it is unpublished.

PROBLEMS

1. Preprints that arrive after the journal version has appeared in print.
2. Same preprint arriving from several different places. The system can track them so they all appear on the appropriate institutional list, but we have to identify them or we end up with multiple entries for the same preprint.
3. Multiple preprints under a single cover – if you file by author, you either take them apart or risk users being unable to find any but the first paper.
4. Abstracts or poster papers that appear as preprints: STScI had 45 and NRAO 50 preprints in the database from IAU Symposium 129 – most were 2 pages or under.
5. People who use the STEP or RAPsheets as a finding guide for what to rip off.
6. Finding citations for things that appear in journals or proceedings we do not receive.
7. Staff who decide every six or eight months to clean their office of accumulated preprints; since we have no idea how long they have had the paper, it means a lot of work checking things against journal tables of contents to avoid entering published papers into the database as unpublished.
8. Not enough hours in the day.