VLB ARRAY MEMO No. 568

(860815)

CSIRO Division of Radiophysics

To: VLBA Memo Series

From: Martin Ewing 7 August 1986

Subj: Relational Databases and VLBA Operations
or, Is the VLBA ready for MIS?

Introduction

I have been looking into the VAX/VMS Rdb (relational database) and Datatrieve (interactive database interface) since they are available here on our VAX 11/750, and since there is some question as to whether NRAO should purchase either or both of these products for the VLBA.

There are a number of issues here, including

(1) What is the need for database systems in a system like the VLBA? How would they be used?

(2) Are VAX Rdb and Datatrieve "good" products, suitable for running on the VLBA AOC computer, for instance?

I found question (1) to be rather interesting, even though (2) is the one more people have been asking. This memo will outline my thoughts on how one could use database systems in the VLBA. I will not go into particular software realizations of these databases here. Some of them, in fact, may remain in their traditional "software" form – pencil and paper.

The Importance of Good Relations

A "relation" is simply a table of "fields" (columns) and "records" (rows) expressing relationships between data. The value of the relational approach (and any other database method) is strongly dependent on the formulation of the relations that are significant in a particular project. In general, one wishes to keep the relations small, and "normalize" (factor) the overall problem as much as possible.

I set out to draw a framework of relations that would encompass the activities of the VLBA. The result is far from a finished product; it has rough edges and is incomplete. Nevertheless, the exercise is instructive, defining how the VLBA can take advantage of modern database technology.

A consistent and general database system will logically extend throughout the VLBA organization, and indeed throughout NRAO. The system naturally, but perhaps

unexpectedly, extends to the whole structure of activity of the Observatory including the telescope proposal process, reviewing, scheduling, observing, and even the publication results. Along the way, personnel issues are touched upon, particularly those involving operations and maintenance.

It is easier to lay out a structure on paper than to implement it in a database system. All of these "relations" must now exist within NRAO, but I suppose they are on various filing systems, sometimes including computers. Whether or not it is sensible to consider placing all of them into a formal database environment, I think it is entertaining, and possibly useful, to carry out the thought experiment.

Some Possible Relations for the VLBA Database

A. Scientific Data Flow

A useful way of looking at VLBA activity is in terms of scientific data flow. This begins with proposals from the investigators (or, perhaps, an "announcement of opportunity" from the Observatory, or an NSF budget,...) and is completed with the publication of the results. In the middle come all sorts of detailed operational data – array schedules, logs, etc. The following database relations may be discerned. They are discussed briefly here and in more detail in an appendix.

1. **OBSERVING PROPOSALS** – Proposals come in, are assigned identification numbers, etc. The database contains the investigators' names, proposal title, reviewer scores, etc.

2. **APPROVED OBSERVING PROGRAMS** – When a proposal is approved, a program number is assigned. The database specifies the resources required for the observation and any significant constraints that must be considered by the schedulers.

3. **ARRAY INSTRUCTIONS** – The schedulers establish this relation, which is a high-level set of instructions for observing, specifying sources, times, frequencies, etc. A fall-back schedule may be included, also.

4. **GENERAL OBSERVING LOGS** – The observing activity generates a log of what actually has happened at the antennas. This is a high-level log, specifying sources, times, program IDs, operator names, and other information that would (in olden days) appear on log sheets.

5. **REAL TIME FRINGE CHECK LOGS** – Real time fringe checks produce a database which is a time series of clock and LO offsets by antenna or baseline.

6. **SIGNAL TAPE LOGS** – This is a brief list of signal tape numbers, and its associated start/stop times. Some data quality information may be included. See SIGNAL TAPE INDEX below.

7. **CORRELATOR INSTRUCTIONS** – Correlation as an activity in the VLBA is similar to observing. It requires schedules and logs to allow smooth operations. The correlator instructions specify what tapes will be required, what correlator modes, etc.

8. **CORRELATOR LOGS** – The correlator log relation contains the record of what programs and what sources were actually correlated, what archive and distribution tapes were generated, and perhaps some data quality measures (dropouts, fringe amplitudes, rates, etc.)

9. **ARCHIVE INDEX** – The archive media must be carefully cataloged so that particular programs and sources can be retrieved. The index will also contain media quality tracking data: purchase date, manufacturer, etc.

10. **DISTRIBUTION TAPE INDEX** – There is presumably a "user" tape library consisting of distribution tapes and other tapes used in postprocessing. This index contains information on tape number, shelf location, ownership, etc.

11. **POSTPROCESSING LOGS** – It may be worthwhile tracking the computing and other data reduction resources used in the course of analyzing data. At very least this could be a computer billing record. Very likely it would be worthwhile to record major mapping runs by investigator, program number, etc.

12. **BASELINES** – This is a list of current list of best solutions to the baselines among all antennas, to be used for correlator control, postprocessing, etc.

13. **PUBLICATIONS** – A publications database relation should contain the bibliographic references generated by each observing program. This may be integrated with a wider library database. Perhaps press releases and other PR material should be included.

B. Resources

14. **SOURCE CATALOG** – The VLBA should maintain its own source catalog relation. In large part, it will contain references to other catalogs, but some specialized information is needed. E.g., is this an official VLBA calibrator? What VLBI flux, structure or visibility is known? Candidate sources or survey areas should also be assigned (temporary?) catalog entries.

15. **ANTENNAS** – Each antenna has its own peculiar data. This would include operational status (e.g., certain receivers or recording modes not working), telephone numbers, and shipping addresses. More data might be required for non-VLBA "foreign" antennas.

16. **SIGNAL TAPE INVENTORY** – The signal tapes are a major resource of the VLBA, and must be tracked and accounted for carefully. The tape log indicates tapes' ID numbers, purchase information, usage history, recording and shipping status, etc. It is related to the SIGNAL TAPE LOG above, which is much smaller. (The two could be merged if desired.)

17. **BANDS SUPPORTED** – This is a list of actual capabilities, including frequency ranges, system temperatures, etc. Might be merged with "ANTENNAS," above.

C. Reference Information

18. **FREQUENCY BANDS** – A list of the official band designators and the corresponding RF frequencies, perhaps including radio astronomy allocations.

19. **RFI STATUS** – A list of known or recurring RFI sources significant for VLBA operations.

20. **DOCUMENTATION (INTERNAL)** – The VLBA memo series (!) and other specifications and documents of interest to VLBA users, engineers, or administrators, including series numbers, authors, titles, keywords, etc.

D. Operations

21. **ANTENNA MONITOR POINTS** – A somewhat summarized database comprising measurements taken at observing time at the antennas. A suitable time scale must be defined (e.g. 1 week?). This database should contain only information that must be retained for longer than this period. More dynamic data, which would be much more voluminous, would presumeably be handled in specialized programs.

22. **ANTENNA SUBSYSTEMS** – This is a list of subsystems at each antenna (or in inventory) with serial numbers, etc.

23. **ANTENNA MODULES** – This relation contains the list of all modules (serial numbers, etc.) presently operating in antenna subsystems.

24. **MEASUREMENT POINTS** – This relation specifies what monitor points are provided for each module, what the meaning of a measurement is (mA crystal current, etc), and what the "red/yellow/green" ranges of measurements are (i.e., when to set alarms).

25. **MODULE INVENTORY** – An inventory of all modules, by serial number, date of manufacture, maintenance history, current status, etc.

26. **MODULE MAINTENANCE** – A record of significant maintenance activities of the Array, including modules repaired, bearings greased, engineers dispatched to the field, etc.

27. **WEATHER LOGS** – A summary of weather information from the antenna monitoring, this would comprise 4X (?) daily readings useful for compiling statistics, etc.

E. Personnel

These data relations can be merged with other Observatory files, and, since they can contain sensitive information, they should be maintained in a secure manner.

28. **USERS** – Identification numbers, names, institutions, past observing records, etc.

29. **STAFF** – Staff ID numbers, names, addresses, positions, and other management information.

30. **REVIEWERS** – A list of reviewers' names, addresses, area of specialization, previous reviewing history, current assignments, etc.

Appendix

1. OBSERVING PROPOSALS
Proposal ID
Title
Principal Investigator User ID
Co-I user ID 1
...
Co-I user ID N
Date of Receipt
Resources Required ... (Time, Freq, No. Ants)
Reviewer ID 1
Date Sent To Rev. 1
Date Rcd from Rev. 1
Rating, Rev. 1
...
Reviewer ID N
Date Sent To Rev. N
Date Rcd from Rev. N
Rating, Rev. N
Overall Rating
Accepted? (T/F)
PI Notification Date
Priority Assigned

2. APPROVED OBSERVING PROGRAMS
Program ID
Proposal ID
Approved Date
Expiration Date
(required antennas)
(required frequencies)
(required times)

3. ARRAY INSTRUCTIONS
(to be issued not too often, e.g. every 5-10 mins.)
Instruction No.
Effective Date/Time
Program ID (special ID for calibrations, real-time fringe checks, etc.)
Source ID
Coordinate Type (cataloged by Source ID, RA/DEC J2000, Galactic, Solar, etc.)
coordinate 1
offset 1
coordinate 2
offset 2

Use Antenna ID 1 (up to 20? antennas per instruction)
...
Use Antenna ID N
LO Freq 1
LO Freq 2
...
LO Freq N
IF Mode
Formatter Mode
Recorder Mode
Receiver Mode

4. GENERAL OBSERVING LOGS
Program ID
Observation (Scan) Number (assigned by operator at obs. time)
Start Date/Time
Stop Date/Time
Success Level (0=nothing, 100=all OK)
Operator Comment

5. REAL TIME FRINGE CHECK LOGS
Source ID
Time of Measurement
Baseline ID
Amplitude
Fringe Rate
Phase

6. SIGNAL TAPE LOGS
Tape ID Number
Start Date/Time
Stop Date/Time
Data Quality Level at Last Correlation

7. CORRELATOR INSTRUCTIONS
Scan ID
Correlation Date/Time
(Correlator Mode)

8. CORRELATOR LOGS
Correlation ID
Scan ID
Correlation Begun Date/Time
Correlation Finished Date/Time
(Antennas Correlated)
(Correlator Modes)
Archive Volume ID
Distribution Volume ID
Operator Staff ID
Success Value
Operator Comment

9. ARCHIVE INDEX
- Archive Volume ID
- Manufacturer
- Product Number
- Acquisition Date
- First Use Date/Time
- Filled Date/Time
- Shelf Location

10. DISTRIBUTION TAPE INDEX
- Distr Tape ID
- Owner
- Shelf position
- Assignment Date
- Expiration Date

11. POSTPROCESSING LOGS
- Post Processing ID
- Correlation ID
- Scan ID
- Program ID
- Postprocessing Run Date/Time
- Scientist Staff ID
- User ID (if present?)
- Computer System ID
- Computer "Billing"
- Processing output Volume ID

12. BASELINES
- Baseline ID (A1-JB1)
- Antenna ID (1) (VLBA1)
- Antenna ID (2) (JBK1)
- Length X (nnnn meters)
- Length Y
- Length Z
- Document ID (best calibr. measurements doc.)

13. PUBLICATIONS
- Publication ID
- Author 1
- User ID 1
- ...
- Author N
- User ID N
- Title
- Subj Type (Astron, Elec Eng, PR, ...)
- Keyword 1
- ...
- Keyword N
- Program ID
- Publ Type (Book, Journal, Newspaper, ...)
- Journal
- publisher

volume
number
page start
page end
date

14. SOURCE CATALOG

Source ID	
coord type	(RA/DEC (J2000), galactic, ecliptic, etc.)
coord 1	
coord 2	
Common Name	
Type	(VLBA Calibrator, temporary program source, etc.)
Entry Date	
Expiration Date	
ID of Reference Catalog	(from whence further info)
ID in Reference Catalog	
Definer ID	(esp if temporary source?)

15. ANTENNAS

Antenna ID	
Site ID	(there may be several antennas/site)
Site Name	(VLA, St. Croix, ...)
Latitude	(nominal, not for precise baselines)
Longitude	
Elevation	
Antenna Diameter	
Telephone No. (voice)	
Telephone No. (data)	
Mail Address	
Standard VLBA Control SW? (T/F)	
Person in Charge Staff ID	

16. SIGNAL TAPE INVENTORY

Tape ID Number	
Manufacturer	
Product Number	
Length	
Acquisition Date	
No. of shipping cycles	
No. of recorder passes	
Date Last Shipped:	
Shipment Originating Site ID	
Last Shipment ID number	
Date Last Received	
Shipment Receiving Site ID	
Present Data Condition	(erased, recorded, released)
Current Site ID	(Correlator, field, or in transit)

17. BANDS SUPPORTED

Antenna ID	
Band ID	
Actual Min Freq	

Actual Max Freq
Nom. System Temp.
Nom. Antenna Efficiency
RCP Available (T/F)
LCP Available (T/F)

18. FREQUENCY BANDS

Band ID	("1.5 GHz")
Official Min Freq.	(1.35 GHz)
Official Max Freq.	(1.75 GHz)

19. RFI STATUS

Band ID	
Antenna ID	(or Site ID?)
Date	
Seriousness Category	(1-10)
Frequency	
Source	
Azimuth	
Modulation	
Identification	

20. DOCUMENTATION (INTERNAL)

Document ID	
Creation Date	
Revision Date	
Author	
Title	
Major category	(HW, SW, User, Policy, ...)

21. ANTENNA MONITOR POINTS
Antenna ID
Subsystem Type ID
Subsystem Number
Module Type ID
Module Number
Measurement Point ID
Date/Time
Value

22. ANTENNA SUBSYSTEMS
Subsystem Type ID
Name
Number per Antenna
Maintainer Staff ID

23. ANTENNA MODULES
Module Type ID
Name
Subsystem Type ID
Number per Subsystem
Maintainer Staff ID

24. MEASUREMENT POINTS
Measurement Point ID
Module Type ID
Name
Units
Minimum value (RED)
Minimum value (YELLOW)
Minimum value (GREEN)
Maximum value (GREEN)
Maximum value (YELLOW)
Maximum value (RED)

25. MODULE INVENTORY
Module ID
Module Serial Number
Date Acquired
Current Status (awaiting repair, in stock, in service, being shipped)
Current Site ID
Current Antenna ID (if any)
Current Subsystem ID (if any)
Installation Date
Previous Site ID
Previous Antenna ID
Previous Subsystem Type ID
Removal Date
Last Maintenance Job ID ID
Date of next scheduled maintenance (if any)

26. MODULE MAINTENANCE
Maintenance Job ID
Previous Maintenance Job ID (for this module)
Module ID
Module Serial Number
Problem Severity (Failed, Marginal, Routine)
Symptoms
Action Taken
Date out of service
Date restored to service condition

27. WEATHER LOGS
Site ID
Measurement Date/Time
Temperature
Dew Point
Bar. Press.
Precipitation
Wind Speed
Wind Direction
Cloud Cover
Comment

28. USERS
User ID (JBlow)

Name	(Joseph Blow)
Address	
Telephone Number	(+1-804-555-0999)
E-Mail address	(jblow@Janksy.UVA.Edu)

29. STAFF
- Staff ID
- Name
- SSN
- Classification
- Supervisor
- Salary
- Date of Employment
- Normal Site ID
- Office No.
- Office Telephone
- Home address
- Home Tel. No.

30. REVIEWERS
- Reviewer ID
- User ID (if any)
- Name
- Address
- Telephone Number
- E-Mail address
- Date of Entry
- Date of Last Review
- Specialty
- Number of Reviews

Important IdentificationVariables

Proposal ID:	A unique ID assigned to each proposal as received.
Program ID:	ID for the observations associated with a particular observing proposal and principal investigator; the scientific research program. E.g. PI user ID + date.
Source ID:	Real source name in standard format, e.g., 2134+004, or a fictitious or provisional source, e.g. "Candidate 01".
Site ID:	Specifies particular observing site, e.g., "VLA", "OVRO", "IOWA", etc. There may be more than one antenna per site, but there will be only one shipping address, etc.
Antenna ID:	Specifies antenna at a particular site, e.g. "27" at site VLA.
Signal Tape ID:	An identifier assigned to each signal tape in the VLBA library. Possibly large enough to include other identification schemes, e.g. Haystack Mark III.
Staff ID:	Employee number?
User ID:	NRAO User number? Some users are also staff?
Reviewer ID:	A unique code for reviewers, who may or may not be users or staff.
Band ID:	A series of standard identifiers, such as "21 cm", "L Band", or "1.4 GHz" for frequency bands.