A COMPENDIUM OF INFORMATION IN SUPPORT OF AN E2E DATA ARCHIVE SYSTEM

The paragraphs written in the "Comment" style are for the benefit of the person writing the document and should be removed before the document is finalized.

VERSION: DRAFT

28 SEPT, 2001

[JOHN BENSON]

REVISION CHART

This chart contains a history of this document's revisions. The entries below are provided solely for purposes of illustration. Entries should be deleted until the revision they refer to has actually been created.

The document itself should be stored in revision control, and a brief description of each version should be entered in the revision control system. That brief description can be repeated in this section. Revisions do not need to be described elsewhere in the document except inasmuch as they explain the development plan itself.

Version	Primary Author(s)	Description of Version	Date Completed
Draft	John Benson	Initial draft created for distribution and review comments	TBD

CONTENTS

New paragraphs formatted as Heading 1, Heading 2, and Heading 3 will be added to the table automatically. To update this table of contents in Microsoft Word, put the cursor anywhere in the table and press F9. If you want the table to be easy to maintain, do not change it manually.

LIST OF FIGURES

New figures that are given captions using the Caption paragraph style will be added to the table automatically. To update this table of contents in Microsoft Word, put the cursor anywhere in the table and press F9. If you want the table to be easy to maintain, do not change it manually.

This section can be deleted if the document contains no figures or if otherwise desired.

1. INTRODUCTION

Author: John Benson

Revision date: 2001/09/13

1.1 Document Motivation

In this document I will layout requirements for the E2E-EVLA Data Archive based on what we now know. The requirements will certainly be upgraded in the near future as we gather information from the NRAO scientific staff and outside users, and as we gain insight into the problem while prototyping a system on the new Storage Area Network device.

We also include lists of the current contents and indexing database of the VLA and VLBA archives, and the corresponding meta-data in an Aips++ Measurement Set.

The ancillary data produced by the VLA is described as well as any other data file that I stumble across.

In the EVLA Memo No. 15, 'Scientific Requirements for the EVLA Real-time System', edited by J. Benson and F. Owen, Appendix A lists the data that the NRAO scientific users think should be archived. I include it again in this document mainly so we don't forget about what we have been requested to write into the archive.

1.2 Reference Materials

List all the documents and other materials referenced in this document. This section is like the bibliography in a published book.

1.3 Definitions and Acronyms

Provide definitions or references to all the definitions of the special terms and acronyms used within this document.

Name	***Definition***	***URL***
AIPS++	*Astronomical Information Processing System*	*http://aips2.nrao.edu*
ALMA	*Altacama Large Millimeter Array*	*http://www.nrao.edu/alma*
EVLA	*Expanded Very Large Array*	*http://www.nrao.edu/evla*
GBT	*Green Bank Telescope*	*http://www.nrao.edu/gbt*
VLBA	*Very Long Baseline Array*	*http://www.nrao.edu/vlba*

2. PRELIMINARY REQUIREMENTS

I need to work on this section quite a bit more. The scope of this document is changing beneath my feet.

2.1 Archive System Capabilities

- Rapid and easy access to all of the archived data through a StarView–like GUI.
- The archive interface and database must be able to connect visibility and image data with the proper corresponding monitor and ancillary dataset in a manner that is transparent to the user.
- The archive system must be fully accessable over the InterNet.
- Two copies of the archived data will be stored separately.
- A detailed cataloging of the archived data in a database system. It should be possible to retrieve visibility data to a resolution of a single scan.

2.2 What Data Will Be In the Archive Dataset

- Raw visibility data as provided by the correlator(s).
- Calibrated/edited visibility provided by Aips++ pipelines.
- Standard images constructed from the archived visibility by a pipeline.
- Monitor and observing log data.
- Ancillary calibration data.
- The on-line system control scripts, maybe 'observing blocks'.
- The Glish control scripts used by the Aips++ pipelines.

2.3 What Information Will Be In the Archive Database.

- Catalog tables for the other archive data types : monitor, logs, ancillary data, control scripts ….
- We will build a catalog or index table that locates scientific archive data at a scan by scan resolution.

2.4 History and Accountability

- **History :** We need to keep records of what operations have taken place on the various archived data. Something like the AIPS History and Log files where the processes that did something to the data are listed, along with the process control parameter values that were used.

- **Accountability :** Accountability is different than history records. Accountablility is achieved generally by tables of numerical values that can be applied to a set of data to back out of a previous operation. Accountability tables contain no information about the process algorithm itself. It is merely a way to back out of changes made to data. Experience with the VLBA has shown that accountability is very important.

2.5 Archive Requirements Due To Pipelines.

2.6 Who or What Will Access the Archive.

3. THE ORGANIZATION AND CONTENTS OF THE CURRENT VLA ARCHIVE

3.1 Astronomy Data Archive

The VLA Archive data is currently stored on nn ExaByte tapes. The archive format is described in the VLA Computer Memorandum 118, 'VLA Archive Data Format', by G.C. Hunt and K.P. Sowinski.

3.2 Meta-Data Catalog Tables

A catalog table in the Ingres database oort::archive, table = sumlist, contains locations and descriptions of observing scans on the archive tapes. The 'sumlist' table is filled from the export format data files contained on the archive tapes. The 'sumlist' table servers as the meta-data catalog table that directs seach queries to the appropriate archive volume and files.

Contents of the VLA Archive Catalog Table :

Parameter	Precision	Comments
apoptions	CHAR*1	
antennae	INTEGER*4	Number of antennae
arrayconfig	CHAR*4	VLA Array Config., A,B,C,D
bandwidth1	FLOAT*8	Bandwidth, MHz
bandwidth2	FLOAT*8	Bandwidth, MHz
bandwidth3	FLOAT*8	Bandwidth, MHz
bandwidth4	FLOAT*8	Bandwidth, MHz
calcode	CHAR*2	VLA Calibration Code
centerofbw1	FLOAT*8	BW Center, MHz
centerofbw2	FLOAT*8	BW Center, MHz
centerofbw3	FLOAT*8	BW Center, MHz
centerofbw4	FLOAT*8	BW Center, MHz
chansep1	FLOAT*8	
chansep2	FLOAT*8	
chansep3	FLOAT*8	
chansep4	FLOAT*8	
correlator	CHAR*4	Correlator Mode
dataselect1	INTEGER*4	

dataselect2	INTEGER*4	
dataselect3	INTEGER*4	
dataselect4	INTEGER*4	
dec1950	FLOAT*8	Dec, B1950
dec2000	FLOAT*8	Dec, J2000
epoch	CHAR*4	Celestial Frame, 2000
filenumber	INTEGER*4	Archive Tape, file number
inttime	FLOAT*8	Integration Time, seconds
mjad	DATE	Obs. Date, Ingres string
netsideband1	CHAR*1	Net sideband, U or L
netsideband2	CHAR*1	Net sideband, U or L
netsideband3	CHAR*1	Net sideband, U or L
netsideband4	CHAR*1	Net sideband, U or L
newtapenum	CHAR*7	New Archive Tape, VSN
numchan1	INTEGER*4	Number Spectral Chans.
numchan2	INTEGER*4	Number Spectral Chans.
numchan3	INTEGER*4	Number Spectral Chans.
numchan4	INTEGER*4	Number Spectral Chans.
observerid	INTEGER*4	VLA Observer Number
observernam	CHAR*1	Program Observer name
obsmode	CHAR*2	
oldtapenum	CHAR*6	Old Archive Tape, VSN
planetary	CHAR*1	Solar System Body, ?
programid	CHAR*6	VLA Observing Code
qualifier	CHAR*1	
ra1950	FLOAT*8	RA, B1950
ra2000	FLOAT*8	RA, J2000
restfreq1	FLOAT*8	Contains 0.0
restfreq2	FLOAT*8	Contains 0.0
restfreq3	FLOAT*8	Contains 0.0
restfreq4	FLOAT*8	Contains 0.0

sourcename	CHAR*16	Source name, users choice
starttimeiat	FLOAT*8	Decimal hours
starttimelst	FLOAT*8	Decimal hours
stoptimeiat	FLOAT*8	Decimal hours
stoptimelst	FLOAT*8	Decimal hours
subarrayid	CHAR*1	Subbary number, def = 1
timeonsource	FLOAT*8	Seconds
velocity1	FLOAT*8	
velocity2	FLOAT*8	
velocity3	FLOAT*8	
velocity4	FLOAT*8	
velsyscode	CHAR*8	

3.3 Monitor and Ancillary Data

Based on conversations with Ken Sowinski and a few other people, what follows is what we think exists in the way of VLA ancillary data.

- Monitor data : contains many measured items all time stamped, probably difficult to decipher the format. Written on 1200 9 track tapes readable on the ModComps. Most useful data may be antenna encoder positions, sync. Detector voltages.
- Pointing data : the measured data from the NRAO pointing runs are stored on the archive tapes. The pointing results are solved for by the VLA operators and placed where?
- Tipping data : the measured data is where? The tipping scan results are solved for by the VLA operators and placed where? In the archive?
- Observe card files : after 1998 – on tape somewhere.
- VLBI Cal. Logs : Only during VLBI observing runs. End up on Aspen in the observing project subdirectory in file xxxxx.ycal.
- Water Vapor Radio. : unknown
- Atmos. Phase Interferometer : unknown
- GPS Receiver data : unknown
- RFI Monitor : unknown
- Surface Weather Station data : unknown

4. THE ORGANIZATIONS AND CONTENTS OF THE CURRENT VLBA ARCHIVE

4.1 Astronomy Archive Data

The VLBA Data Archive is contained in uv_FITS files written and stored on 4000 DAT tapes. The uv_FITS file format is described in AIPS Memo No. 102, 'The FITS Interferometry Data Interchange Format', by Chris Flatters.

There is only one copy of each tape. A tape may contains files from unrelated observing programs. Monitor and logging data is stored in computer diskfiles initially and later on backed-up to tape.

4.2 Meta-Data Catalog Tables

The VLBA Archive is cataloged in a table = 'fxarchive' in database = oort::vlba_mon. The description is pretty minimal, mainly relating observing project id's (proposals) to archive tape vsn's and file numbers.

Contents of the VLBA Archive Catalog Table :

Parameter	Precision	Comments
Arch_vsn	CHAR	Archive Tape VSN
Arch_fileseq	INTEGER*4	Archive Tape File Number
Filename	CHAR	VLBA Correlator Filename
Proposal	CHAR	Observing Proposal Id
Segment	CHAR	Observing Proposal Segment
Jobnum	INTEGER*4	Correlator Job Number
Arch_date	DATE	Archive Date, Ingres string
filesize	INTEGER*4	File size in bytes
Arch_drive	CHAR	Written on this tape drive

4.3 Monitor and Ancillary Data

5. VLA META-DATA FROM AN AIPS++ MEASUREMENT SET

Describe the measurement set, etc….

The Aips++ Measurement Set Tables are abbreviated as follows :

- AN : antenna table
- FL : field table
- HI : history table
- MN : main table
- OB : observation table
- PL : polarization table
- PR : processor table
- SW : spectral_window table

Parameter	**Cell Value**	**Aips++ MS Parameter**	**Cell Value**
apoptions			
antennae	26	MN:antenna1, antenna2	
arrayconfig	A		
bandwidth1	50.0	SW(1):total_bandwidth	50.00e+6
bandwidth2	50.0	SW(2):total_bandwidth	50.00e+6
bandwidth3	50.0	SW(1):total_bandwidth	50.00e+6
bandwidth4	50.0	SW(2):total_bandwidth	50.00e+6
calcode		FL(6):code	C
centerofbw1	8435.100	SW(1):ref_frequency	8.435100e+9
centerofbw2	8485.100	SW(2):ref_frequency	8.485100e+9
centerofbw3	8435.100	SW(1):ref_frequency	8.435100e+9
centerofbw4	8485.100	SW(2):ref_frequency	8.485100e+9
chansep1	0.00	SW(1).chan_width	50.0e+6
chansep2	0.00	SW(2).chan_width	50.0e+6
chansep3	0.00	SW(1):chan_width	50.0e+6
chansep4	0.00	SW(2):chan_width	50.0e+6
correlator		PR:type, sub_type	‘unavailiable’

dataselect1	0		
dataselect2	0		
dataselect3	0		
dataselect4	0		
dec1950	-0.385	none	
dec2000	-0.390	FL(5):phase_dir(2)	-0.389650481
epoch	2000	FL(5):phase_dir	col. keyword ref = J2000
filenumber	7		
inttime	10.0	MN:interval	10.0
mjad	14-nov-2000	MN:time	
netsideban1	L	SW(1):net_sideband	1
netsideban2	L	SW(2):net_sideband	1
netsideban3	L	SW(1):net_sideband	1
netsideban4	L	SW(2):net_sideband	1
newtapenum	XH00149		
numchan1	0	SW(1):num_chan	1
numchan2	0	SW(2):num_chan	1
numchan3	0	SW(1):num_chan	1
numchan4	0	SW(2):num_chan	1
observerid	266		
observernam	BENSON J.	OBS:observer	‘unavailiable’
obsmode			
oldtapenum	N13755		
planetary	0		
programid	AB973	OBS:project	AB973
qualifier	0		
ra1950	3.387	none	
ra2000	3.399	FL(5):phase_dir(1)	-2.88455822
restfreq1	0.000		
restfreq2	0.000		
restfreq3	0.000		

restfreq4	0.000		
sourcename	1258-223	FL(5):name	1258-223
starttimeiat	1.385	MN:time	
starttimelst	2.327	MN:time	
stoptimeiat	1.515	MN:time	
stoptimelst	2.457	MN:time	
subarrayid	1	MN:array_id	
timeonsoure	1400.000	MN:time	
velocity1	100.000		
velocity2	200.000		
velocity3	100.000		
velocity4	200.000		
velsyscode	TF FTF F		

6. SCIENTIFICALLY RELEVANT DATA TO BE ARCHIVED

6.1 Visibility Data

- date and time of record
- integration time
- baseline_id
- pointing center
- phase-center
- freq_id
- subarray_id
- channelization description
- describe on-line calibration already applied
- visibility data weights
- u, v, w coordinates
- raw correlator output visibilities

6.2 Source Table

- source type (star, planet, tipping, phase-center)
- source IAU name
- RA and DEC
- RA and DEC first derivatives
- coordinate reference frame
- position and rate reference epoch
- distance (parallax)
- calibrator source type
- label (text string from OBSERVE control script)

6.3 Antenna Location Table

- station name (VLBA_KP for VLBA, AW32 for VLA, NMA_BN for NMA)
- antenna_id
- pad number
- x, y, z (geocentric position)

- axis type
- axis offset

6.4 Antenna Table

- antenna_id
- start_time, stop_time
- subarray_id
- pointing offsets (both in ra/dec and az/el)
- LO_offsets (one for each baseband)
- LO_phase_offsets (one for each baseband - could support auto-phasing
- this way)
- fast_integration_flag (different integration times per antenna)

6.5 Subarray Table

- subarray_id
- observing project_id
- correlation_configuration
- LO_configuration
- source_id (one for each baseband or model_id)

6.6 Frequency Table

- antenna_id
- date and time of reconfiguration
- sky frequency of 4 pairs of 2 GHz baseband channels
- net sideband for each 2 GHz baseband channel
- polarization for each 2 GHz baseband channel

6.7 Local Oscillator Offset Table

- antenna_id
- date and time of reconfiguration
- LO offset for each of 4 pairs of 2 GHz baseband channels

6.8 Correlator Table

- subarray_id
- date and time of reconfiguration
- for each of 16 subbands in each 2 GHz baseband channel :
    - subband lower edge frequency
    - subband bandwidth
    - net sideband
    - n_lags for subband correlator
    - long-term accumulator integration time
    - range(s) of spectral pixels written out
    - polarization mode : cross-hands on/off

6.9 Auto-phasing Table

- antenna_id
- date and time of reconfiguration
- phase correction applied to each 2 GHz baseband channel
- quality measure of the auto-phasing

6.10 Interferometer Model Accountability

- date and time of model
- time interval
- antenna_id
- source_id
- total group delay, nth order polynomial (8 possible phase-centers)
- total phase delay, nth order polynomial (8 possible phase-centers x 16 subbands)

6.11 Flags Table

- date and time range - flag condition active
- antenna_id
- baseband bad flag (8)
- subband bad flag (8,16)
- polarization bad flag (4)
- RFI flagging description

- reason for flag

6.12 Antenna Pointing Corrections

- antenna_id
- date and time of measurement
- pad number
- pointing mode description
- date and time of measurement
- date and time of application
- delta azimuth, delta elevation

6.13 Antenna Polarization Corrections

- antenna_id
- date and time of measurement
- frequency dependence
- feed(s) position angles
- polarization correction parameters

6.14 Antenna Focus

- antenna_id
- date and time of measurement
- receiver band - need more
- subreflector position correction
- subreflector angle correction

6.15 Antenna Pointing Vectors

- antenna_id
- date and time of encoder readout
- azimuth, elevation angles

6.16 Pulsar Gate Weights

- subarray_id
- date and time of correlator integration interval
- gate weights for each gate bin per subband

6.17 System Temperatures

- date and time of measurement
- antenna_id
- freq_id
- baseband channel
- system temperature

6.18 Total Power Measurements

- date and time of measurement
- antenna_id
- freq_id
- baseband channel
- system temperature (probably one for each subband)

6.19 Atmospheric Phase Interferometer

- date and time of measurement
- amplitude and phase : 1 second intervals
- RMS and monitor points : 1 minute intervals

6.20 Surface Weather Measurements

- date and time of measurement
- weather station id
- barometric pressure, temperature, dew point, wind speed and direction
- precipitation measure

6.21 WVR Data

- date and time of measurement
- antenna_id
- WVR frequency bands
- raw WVR output for each WVR frequency band
- structure function fitting data products
- phase correction (one for each subband)

6.22 GPS Data - not yet specified

6.23 RFI Data - not yet specified

7.

8.

Include supporting detail that would be too distracting to include in the main body of the document.