

## 10. DATA MANAGEMENT

---

*Gustaaf van Moorsel*  
*Last changed 2003-Aug-25*

### Revision history

2001-July-18: Initial version 1.0 by Tim Cornwell

2003-August-25: Version 1.1 by Gustaaf van Moorsel

### 10.1 Summary

Originally, Data Management deliverables for EVLA were planned to be built as part of the NRAO-wide End-to-End project (e2e). The recent reorganization of e2e management, however, calls for deliverables which are primarily focused on the EVLA. Though we do not plan to lose sight of a potential observatory-wide use of the various tools, the majority of our current resources will be devoted to EVLA specific deliverables. The most recent version of the e2e Project Book (<http://www.nrao.edu/e2e/documents/e2eprojectbook.doc>), though reflecting the previous, more NRAO-wide oriented, approach, continues to be a fair representation of the more restrictive, EVLA specific, case.

### 10.2 Overview of Deliverables

For the EVLA, the software to be delivered must be able to handle:

1. Proposal Preparation, Submission and Handling
2. Observation Preparation
3. Observation Scheduling
4. Data Archive
5. Image Pipeline
6. Data Post-processing

These are to be treated as a whole, so that information flows through the above steps without loss due to e.g. format changes, omissions, etc. In addition, each of these deliverables must interact closely and seamlessly with the Monitor & Control (M&C) effort which is covered in Chapter 9. For some deliverables, such as Proposal Preparation, interaction with M&C will be modest; for others, such as Observation Scheduling and Data Archiving, interaction will be frequent and intense. A more detailed description of each of these deliverables is given in Section 11.4.

### **10.3 Requirements**

During 2002 and much of 2003 a number of e2e requirement documents was written. Each document contains the requirements on the e2e software from a particular point of view, e.g. Scientific, Operations, etc. The currently available list of requirement documents is:

- Operations
- Engineering
- Scientific
- Real-time
- Post-processing

Each requirement is given a priority and a timescale for completion. We currently are in the process of combining all requirement documents after which we will prioritize based on the available resources. Given current resources, it is not unlikely that a certain degree of de-scoping will have to take place.

### **10.4 Description of Deliverables**

The six main categories of e2e deliverables are described below. In the following, we report on their current status, general priority, and timeline. A number of further e2e tasks, which have substantial overlap with M&C and do not clearly belong in any of the main categories, is listed in Section 11.5.

### **10.4.1 Proposal Preparation, Submission, and Handling**

The proposal submission software allows scientists to submit proposals to the EVLA, ensuring that the information necessary to the review of the proposal is included and, as much as possible, is correct. An observer will use the proposal submission software to construct, verify, and submit a proposal for the EVLA. This may be done collaboratively with other scientists at other locations

It does not require detailed descriptions of the observations, nor does it have capabilities for detailed planning of observations (both responsibilities of the Observation Preparation software). However, in many cases, it will be possible to derive a workable description of the observation from the information in the proposal alone. This description may then require substantial subsequent modifications.

Work on an NRAO-wide proposal system has been underway for some time as part of the e2e project; a working prototype exists, and the first round of beta testing was concluded recently. We intend to redefine this effort in terms of EVLA requirements shortly. Since according to the Scientific Requirements Document this deliverable will be due Q2, 2007, we are unlikely to give further development in this area high priority in the near future.

Proposal handling is one area in which we intend to keep an observatory-wide perspective. We realize it is in NRAO's interest for proposal related software to have similar look and feel regardless of instrument, and we plan to coordinate our efforts in this area with other NRAO telescopes. We also intend to familiarize ourselves more with related ongoing ALMA efforts.

### **10.4.2 Observation Preparation**

The observation preparation software allows description of observations to be made with the EVLA, and is a natural successor to existing tools such as observe and JObserve. The key innovations over the existing software will be:

1. It will provide a sufficiently high level description of the observation that the pipeline can derive suitable processing heuristics. Alternatively, the Observation Preparation software will contain instructions to the pipeline.
2. Eventually, it will be able to drive observation planning software

The Scientific Requirements Document lists the needed availability of this tool as Q2, 2007. We intend to formulate a number of intermediate milestones to meet this target date.

The output of the Observation Preparation Software will be Observing Blocks, which form the input to the Scheduling Software. The core of each block consists of an observing script which communicates directly with the antenna control computer. Selection of a language to be used in the scripts is underway; we need such a scripting language by the time we want to control the test antenna in Q4, 2003.

### **10.4.3 Observation Scheduling**

The Observation Scheduling Software allows NRAO telescope operators to schedule a telescope from Observation Descriptions in the form of scheduling blocks created by the Observation Preparation software. Telescope operations staff will receive these scheduling blocks and feed them to a queue for subsequent observation. This queue may be dynamically scheduled or fixed order. The Observation Scheduling Software is not controllable by an observer but the current queue may be viewed by anyone.

Work on telescope scheduling software has already been ongoing as part of the e2e effort. A prototype for implementing dynamic scheduling at the VLA is under development. With EVLA requirements now available, we will redefine this effort into an EVLA one, with anticipated delivery in Q4, 2005.

### **10.4.4 Data Archive**

The archive system allows interactions with the archive, including submission, querying, distribution requests, *etc.*

A telescope control system may submit data to be archived, or a request for some archived data. A scientist may submit a query to a catalog or a specific distribution request.

The archive will have to contain scientific (visibility) data, calibration data, and monitor data. The Scientific Requirements state a Q2 2007 delivery for visibility and calibration data, but engineering/ monitor data will be needed well in advance of that. In particular, monitor data from the test antenna needs to be archived. We currently have the mechanism in place to do so.

Recent e2e archiving work has concentrated on the VLA; we are confident that it can be expanded to include EVLA requirements.

### 10.4.5 Image Pipeline

The Pipeline system allows submission, monitoring, and interaction with a pipeline reduction system in AIPS++.

A pipeline may run in a variety of contexts:

- In a user's desktop machine,
- In a user's personal network,
- In a dedicated telescope-based network,
- In a Grid.

It is important that the pipeline run the same software as that used for interactive data reduction, in this case AIPS++. The pipeline is split conceptually into *mechanism* and *heuristics*. The pipeline heuristics are rules to help in the processing of observations on the pipeline. They are driven by the data itself, pipeline processing parameters, and meta-data. EVLA specific heuristics will have to be developed.

As with the Post-processing software (see below), for the next 12 months we intend most pipeline related development to be focused on ALMA. After this time the EVLA project will be in a position to benefit from the experience gained by ALMA, and start developing its own version, applying its own requirements as set forth in the various requirements documents. The current schedule has Pipeline Data Processing for the EVLA available by Q2 2007.

### 10.4.6 Post-processing Software

Post-processing software will be delivered by the AIPS++ group. Until the middle of 2004, the lion's share of AIPS++ effort will be focused on ALMA acceptance testing in July 2004, and we expect few, if any, EVLA specific applications. ALMA acceptance testing will concentrate on three main aspects: robustness, speed, and completeness. We expect that success in these three areas will benefit the EVLA as much as ALMA. The EVLA will require significant instrument-specific application development which is of lesser importance and interest to ALMA; we expect that this work will start in earnest only after ALMA has adopted AIPS++. These areas include:

- Wide field imaging
- Wide band imaging

- RFI excision

## 10.5 Intersections with Monitor and Control

There are a number of places where Monitor and Control and e2e intersect. Such software elements are hard to categorize as either e2e or M&C. It is in addressing these issues that a structure stimulating closer contact between e2e and M&C can be particularly effective. In other cases, the M&C system may impose a much earlier completion date than the one given in the current requirements documents. A current list of issues:

- Control Scripting Language. This is an important element of the e2e Observation Preparation Software, but will probably have to become the M&C group's responsibility. Its urgent need is dictated by the test antenna schedule: it will be needed for pointing the test antenna some time during Q4 2003.
- Wrapper for the CALC program. CALC is used to provide the interferometer model to the correlator. It will be needed to support pointing of the test antenna which gives it the same time frame as the control scripting language - Q4 2003.
- Real-time Calibrator Analysis Tool (RTCAT). RTCAT accepts the visibility data for calibrators as input and provides complex antenna gains as output. It will be required in a number of important observing modes, such as reference pointing, auto-phasing for VLBI, data quality displays, and determination of focus parameters. If we wish to retire the Modcomps in Q1 or Q2 2005, then a subset of the full RTCAT functionality as specified in the EVLA e2e Science Software Requirements document will be needed by Q3 2004 or early Q4 2004. The next date after Q3 2004 is the need to support the prototype correlator, which is scheduled to arrive in Q4 2005. RTCAT must function as a real-time or very near real-time element. Since it may be impossible to obtain current visibility data from the archive quickly enough, a fast path for getting data to RTCAT and the results from RTCAT to the EVLA M&C system may be needed.
- A possible later reworking of software development for the test antenna. For example, the current approach taken for the test antenna monitor data archive is an experiment. If this experiment is successful, as it appears it will be, this item will not be necessary.
- Visibility Data Archive. According to the Scientific Requirements this is needed by Q2, 2007. Other drivers call for an earlier delivery date: Retiring the Modcomps (implies archiving the output of the current correlator), capturing the output of the prototype correlator (Q4 2005), and archiving the output of the first functional fraction of the WIDAR correlator (Q4 2006).

- Quick-look pipeline tool. One requirement of the EVLA e2e Science Software Requirements document states “The time-lag between when data is collected by the real-time system and when it is available to the Quick-Look Pipeline Tool should be less than 30 seconds. This may guide a decision on whether data is all archived first, and then retrieved from the archive into the Tool, or fed directly from the real-time system into it”. This requirement may force us to abandon the simple model that visibility data goes only to the archive and that all needs for visibility data are satisfied by the archive.
- Astronomer’s “What’s Up” Display. As for the Quick-look Pipeline tool, the need date for this display is currently driven by the date at which the Modcomps are to be retired.
- Data from the correlator back-end (a M&C effort) will have to be fed in to the science archive (an e2e deliverable). Issues regarding data format will have to be resolved.

## 10.6 References

1. EVLA Array Operations Software Requirements  
(<http://www.aoc.nrao.edu/evla/techdocs/computer/workdocs/array-sw-rqmts.pdf>)
2. EVLA Engineering Software Requirements  
([http://www.aoc.nrao.edu/evla/techdocs/computer/workdocs/engr\\_requirements\\_2003aug06.pdf](http://www.aoc.nrao.edu/evla/techdocs/computer/workdocs/engr_requirements_2003aug06.pdf))
3. EVLA e2e Science Software Requirements  
([http://www.aoc.nrao.edu/evla/techdocs/computer/workdocs/evla\\_requirements\\_15apr03.pdf](http://www.aoc.nrao.edu/evla/techdocs/computer/workdocs/evla_requirements_15apr03.pdf))
4. EVLA Real-time Software Requirements (*in preparation*)
5. EVLA Post-processing Software Requirements (DRAFT)  
([http://www.aoc.nrao.edu/evla/techdocs/computer/workdocs/evla\\_offline.pdf](http://www.aoc.nrao.edu/evla/techdocs/computer/workdocs/evla_offline.pdf))
6. e2e Project Book (<http://www.nrao.edu/e2e/documents/e2eprojectbook.doc>)