The Science Data Model for ALMA and EVLA: The Triumphs and Pitfalls of Software Sharing and Reuse

Nuria Lorente
NRAO
ALMA Computing
• Michel Caillat (ALMA, Observatoire de Paris)
• Robert Lucas (ALMA, ESO)
• Rich Moeser (EVLA, NRAO)
• Martin Pokorny (EVLA, NRAO)
• Michael Rupen (EVLA, NRAO)
• François Viallefond (ALMA, Observatoire de Paris)
• Andreas Wicenec (ALMA, ESO)
Software Sharing / Reuse

• Generally hailed as “Good Practice”
  – Reduces developmental costs
  – Makes use of expertise in the community
  – Avoids reinvention of the wheel

• Broad range of meanings
  – You're welcome to use my code
  – Project developed with reuse and generality in mind
  – Must strike a balance between generality and pragmatism.
Software Sharing / Reuse

We know this, so...

why is the amount of software sharing/reuse in the Astronomical community smaller than it could be?

• Small size of the Astronomical community
  – Projects with similar software needs rarely overlap in time.

• Obtaining political support can be difficult
  – Extra cost to the institution NOW for potential LONG-TERM benefits to the community.
ALMA / EVLA Software Sharing Project

• Unique position of overlap
  – Time: concurrent development
  – Institutional involvement: NRAO is playing a significant role in both projects.

• Development of a common Science Data Model
ALMA

- Atacama Large Millimeter / submillimeter Array
- Partnership of Europe, Japan and North America, in cooperation with the Republic of Chile.
- A 66-antenna array at a high-altitude (5000m) site in northern Chile, operating at λ~0.85-3mm.
- Early science expected in 2011.
EVLA

- Expanded Very Large Array
- Multiply the VLA's capabilities ten-fold
- $\lambda=1\text{-}50\text{GHz}$, $1\mu\text{Jy}$ point-source continuum sensitivity, new correlator with $8\text{GHz}$/polarisation capability.
- Completion by 2012
ALMA / EVLA Science Data Model (SDM)

- The SDM must contain all the information necessary for the astronomical processing of raw data from the telescope.
- Format in which raw science data will be archived and provided to astronomical observers.
- Supported by the Common Astronomy Software Applications (CASA) post-processing software
  - Measurement Set filler
Science Dataset

- Set of tables represented as XML documents
- Data from correlators, square-law detectors, radiometers, etc. stored in binary blocks.
Science Data Model Definition

- Binary Data Format Specification (BDF)
  - Text document
- Binary Data XML Schemata
  - Hand-crafted
- Meta-data Protocol Specification
  - UML & text document
- Meta-data XML Schemata
  - Auto-generated
What have we learned?

• People's work priorities
  – Very easy for the shared project to get lower priority than one's “main” project

• The smaller project's needs are sometimes lost in the momentum of the larger project
  – Becomes software reuse

• Communication
  – Most significant issues were not to do with the “shared” aspects, but by the distributed nature of the shared project
What have we learned?

Timing conflicts between the ALMA and EVLA projects

- Individual project priorities take precedence, which leads to
- Duplication of features
- Sometimes temporary implementations are replaced by the “common version”, but at other times this results in
- Divergence of software, for non-technical reasons.
Conclusion

• Software sharing / reuse
  – Possibility of many benefits
  – To the organisations involved
  – To the wider community

• Makes best use of areas of expertise

• Encourages development of more flexible and general solutions to problems
  – Need to reinvent the wheel less often!
Conclusion

BUT:

• It does not have zero cost
  – Will only happen if effort to reuse/share < effort to rewrite!

• Requires significant support
  – Staff involvement
  – Project priorities and timelines

• Major benefits are likely to be noticed long-term