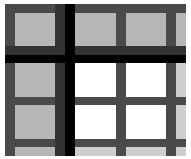
Tabular Data Access For Astronomy (and beyond?)

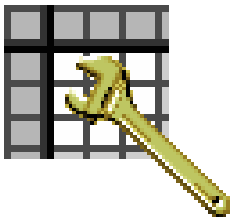
Mark Taylor, Astrophysics, Bristol University



STIL

- Public API for table I/O and processing

  http://www.starlink.ac.uk/stil/



STILTS

- Command-line tools for table manipulation

  http://www.starlink.ac.uk/stilts/

TOPCAT

- GUI table visualisation/manipulation

  http://www.starlink.ac.uk/topcat/

Tables in Astronomy

- Mostly source catalogues
- Many columns (10^1-10^3?)
- Many rows (10^2-10^9?)
- Metadata is important
  - per-table and per-column
- Various formats
- Generally static

Data Integration

- VOTable important new format for e-science era
  - metadata-rich
  - XML-based (pure XML or XML/binary)
  - initially, not much software around
- But many other formats exist
  - FITS, SQL, ASCII, ...
- Other formats won't disappear
  - different formats suitable for different jobs
  - legacy data
  - conservatism
- Take a pessimistic approach
  - people will use many formats – deal with it

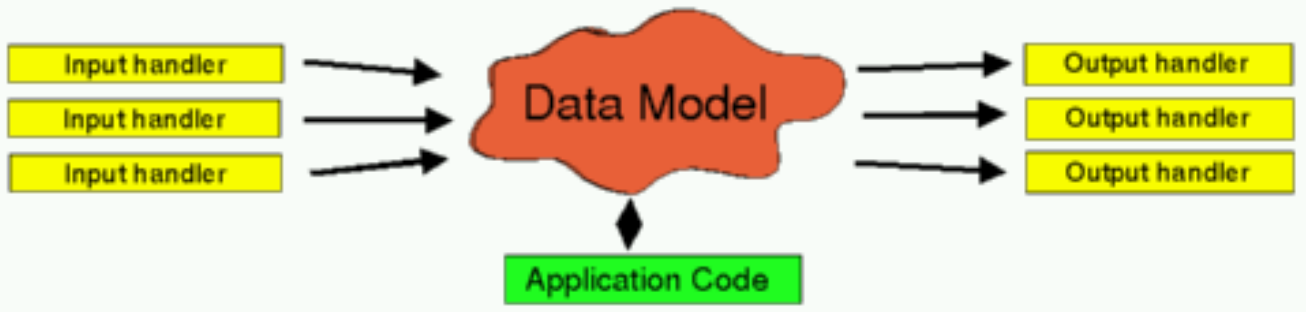
Design Goals

- Multi-format
  - and extensible to new formats
- Scalable
  - streaming, not memory limited
- Easy to deploy

Java

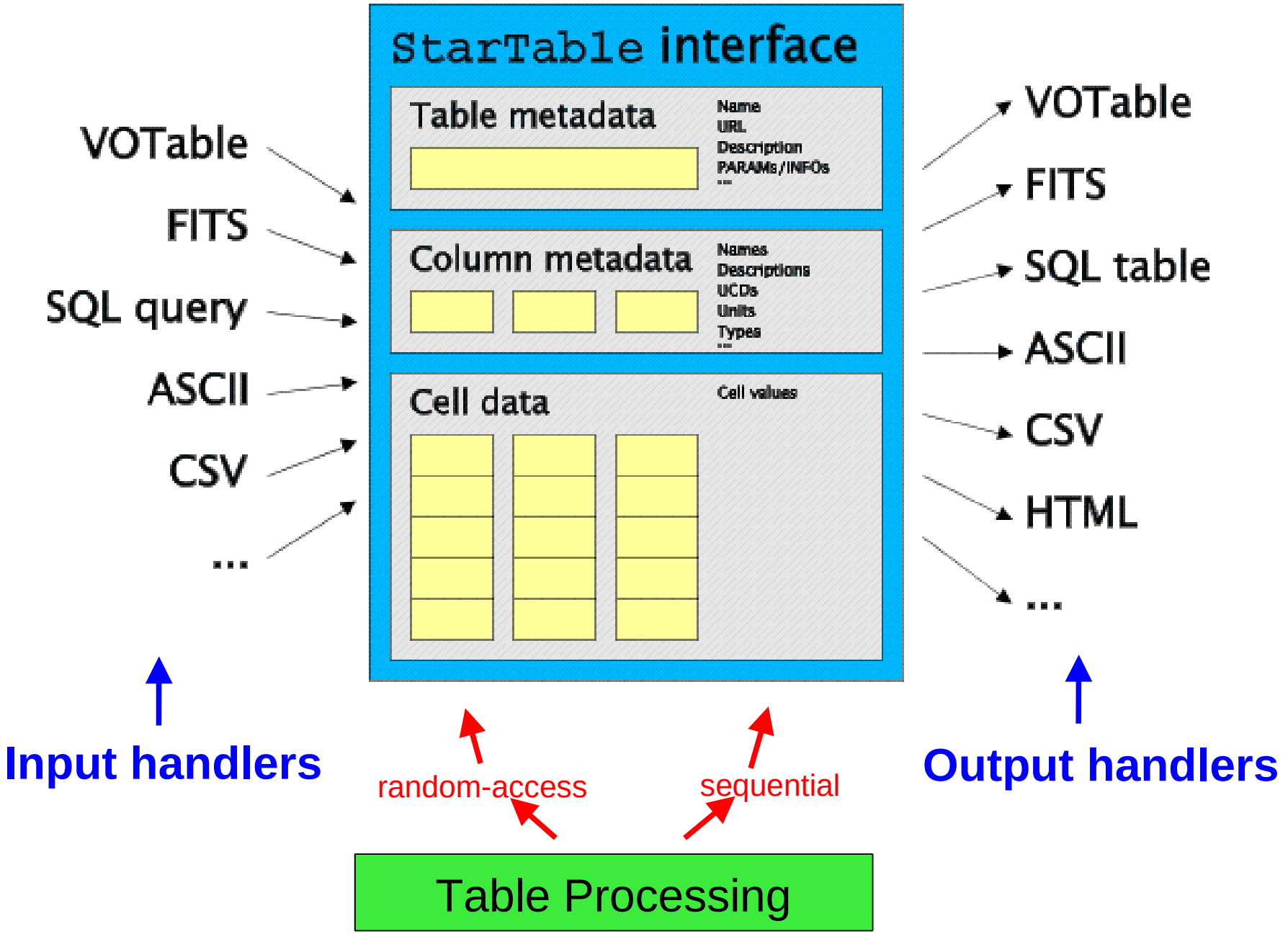
- Very easy to deploy across platforms
  - write once, run anywhere
  - no native libraries (well, almost)
  - webstart
- Many required libraries in place
  - SAX, DOM, JDBC, networking, graphics, ...
- Object Orientation facilitates pluggability
- Performance fine
- Virtual memory handling not very sophisticated

Format-Neutral Architecture



- No native/preferred storage format
- Abstract data model/interface definition instead
- Application code doesn't know what table's format is
- I/O handlers know how to de/serialize
- Comes with a set of standard I/O handlers
  - more can be added later, plugged in at runtime if desired
- Any-to-any format conversion comes for free
- Better get the model right!

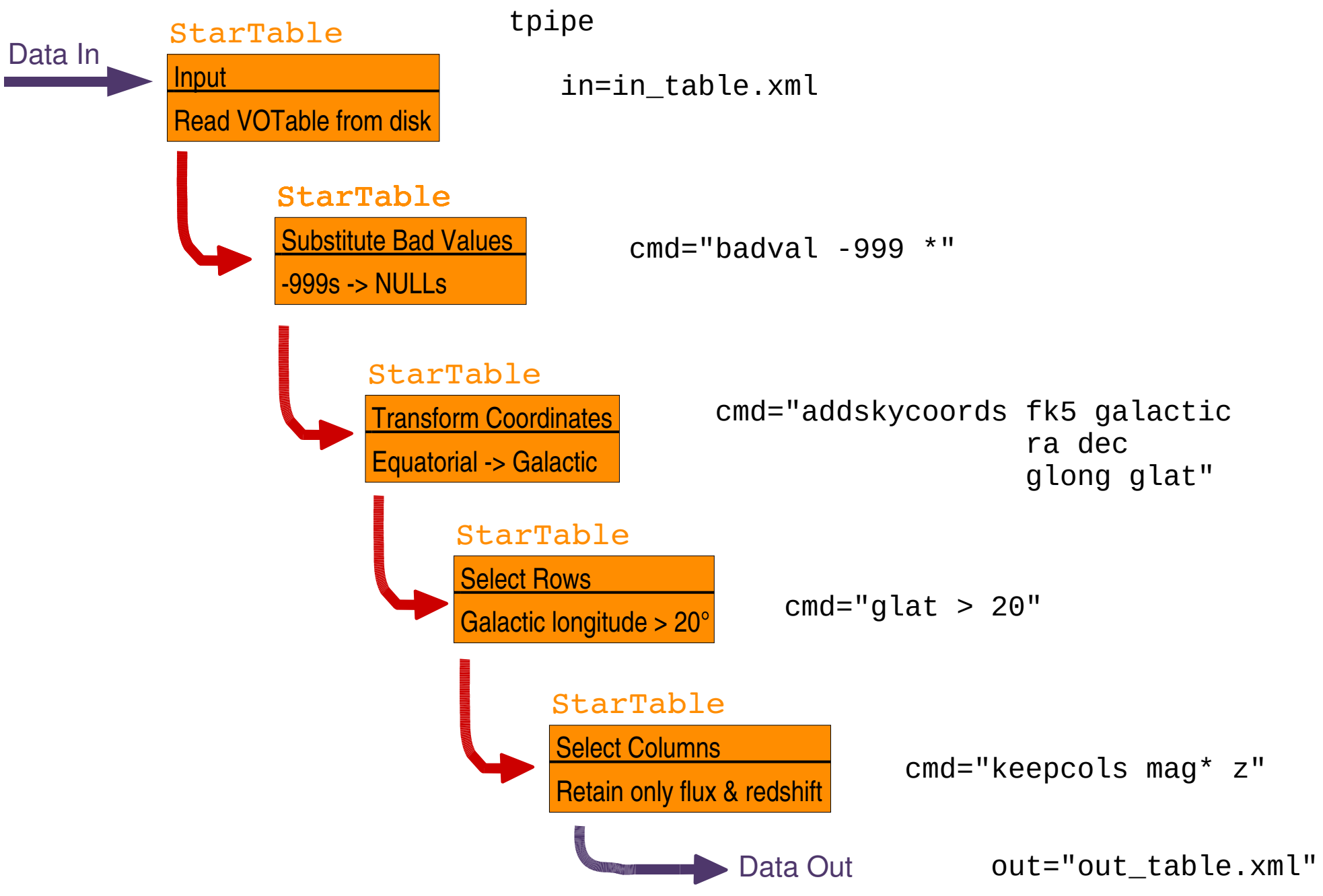
STIL's Data Model



StarTable interface

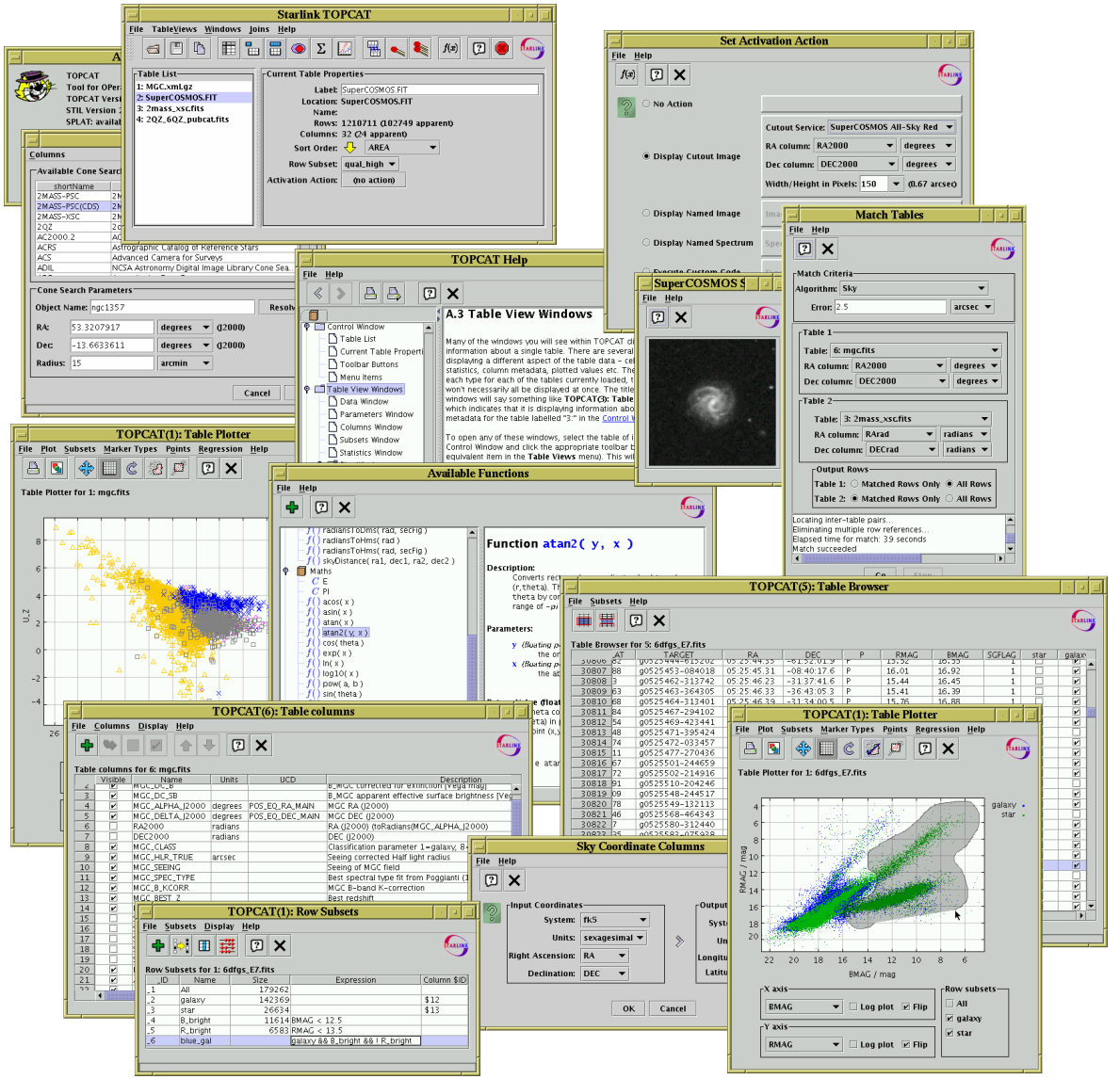
```
public interface StarTable {

    /* Table metadata. */
    URL getURL();
    void setURL( URL url );
    String getName();
    void setName( String name );
    List getParameters();
    DescribedValue getParameterByName( String parname );
    void setParameter( DescribedValue dval );

    /* Column metadata. */
    ColumnInfo getColumnInfo( int icol );
    List getColumnAuxDataInfos();

    /* Table shape. */
    int getColumnCount();
    long getRowCount();

    /* Table data. */
    boolean isRandom();
    RowSequence getRowSequence() throws IOException;
    Object getCell( long irow, int icol ) throws IOException;
    Object[] getRow( long irow ) throws IOException;
}
```

Table Processing



TOPCAT

- Visualisation
  - 2d/3d scatterplots
  - 1d/2d histograms
  - table browser
- Analysis
  - statistics
- Processing
  - sorts
  - selections
  - rearrangements
  - calculations
  - spatial/other joins
- ... and more (demo)



Software Status

- STIL/STILTS/TOPCAT is:
  - available
  - in use (mostly astronomy, some biomedical interest)
  - under continuing development
  - supported
  - fully documented (HTML, PDF, within TOPCAT)
  - open source

Summary

- Format-neutral model good for data integration
- Java good for implementation
- Any takers for non-astro use?

http://www.starlink.ac.uk/topcat/
/stilts/
/stil/