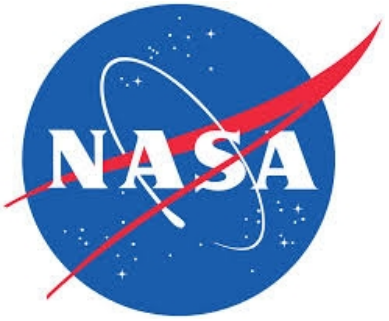
Tables in HDF5

Walter Landry



Tablator
The Universal Table Translator

- Users give us all kinds of files, and want results in all sorts of formats.

Tablator
The Universal Table Translator

- Users give us all kinds of files, and want results in all sorts of formats.
- Reads (with limitations)
  - FITS
  - IPAC Table
  - HDF5
  - VOTable
- Writes
  - FITS
  - IPAC Table
  - HDF5
  - VOTable
  - CSV, TSV
  - HTML

Tablator
The Universal Table Translator

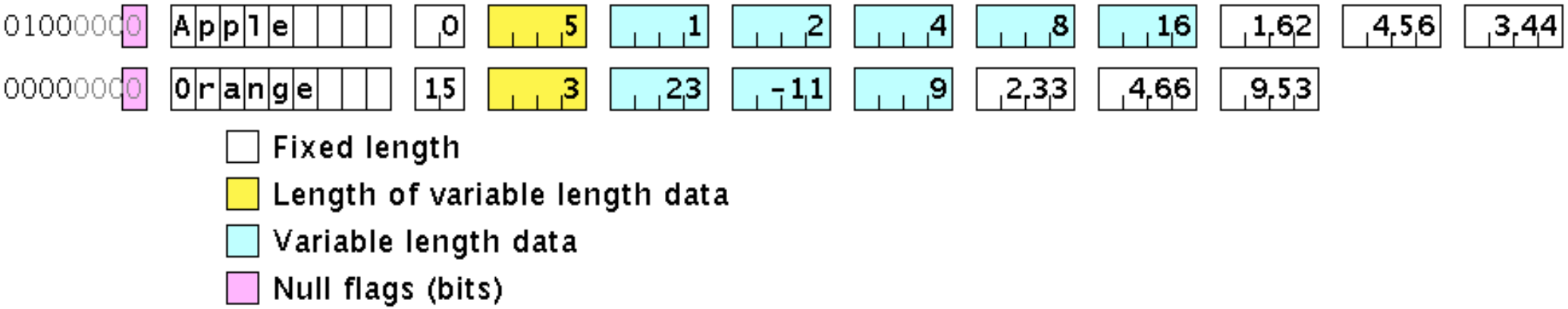
- Users give us all kinds of files, and want results in all sorts of formats.
- Reads (with limitations)
  - FITS
  - IPAC Table
  - HDF5
  - VOTable
- Writes
  - FITS
  - IPAC Table
  - HDF5
  - VOTable
  - CSV, TSV
  - HTML

https://github.com/Caltech-IPAC/tablator

Still a work in progress

Internal Data Structure

- The raw data is VOTable Binary2



- I do not support variable length structures yet.

Internal Column Metadata Structure

- Column metadata
  - Description
  - Attributes (key,value string pairs)
  - Values
    - min, max, ID, null, ref
    - Option
      - key, value
      - value can be an Option
  - Links (URL references)
- Most of this complication is to support VOTables

Internal Non-Column Metadata Structure

- Separate field for description

Internal Non-Column Metadata Structure

- Separate field for description
- Non-column metadata is mostly a list of key-value pairs

Internal Non-Column Metadata Structure

- Separate field for description
- Non-column metadata is mostly a list of key-value pairs
  - key
    - string
    - not unique

Internal Non-Column Metadata Structure

- Separate field for description
- Non-column metadata is mostly a list of key-value pairs
  - key
    - string
    - not unique
  - value
    - string and list of key-value pairs
      - sublist is for xml attributes
      - sublist keys are unique

Internal Metadata Format

- No explicit hierarchy beyond that
- However, if you create a key foo.bar, then serialization into a VOTable creates bar as a sub-element of foo.

Reading HDF5

- HDF5 api really wants a file name, not a stream
- We read the whole file at once

Writing HDF5 Metadata

- I map the metadata to a single HDF5 attribute
- That attribute consists of a list of key-value pairs
- I can not use separate HDF5 attributes for each piece of metadata, because HDF5 attribute names must be unique.
- On my todo list is to make the attributes hierarchical as in VOTable.

Writing HDF5 Column Data

- The columns map to HDF5 types
  - except Bool
  - no duplicate column names

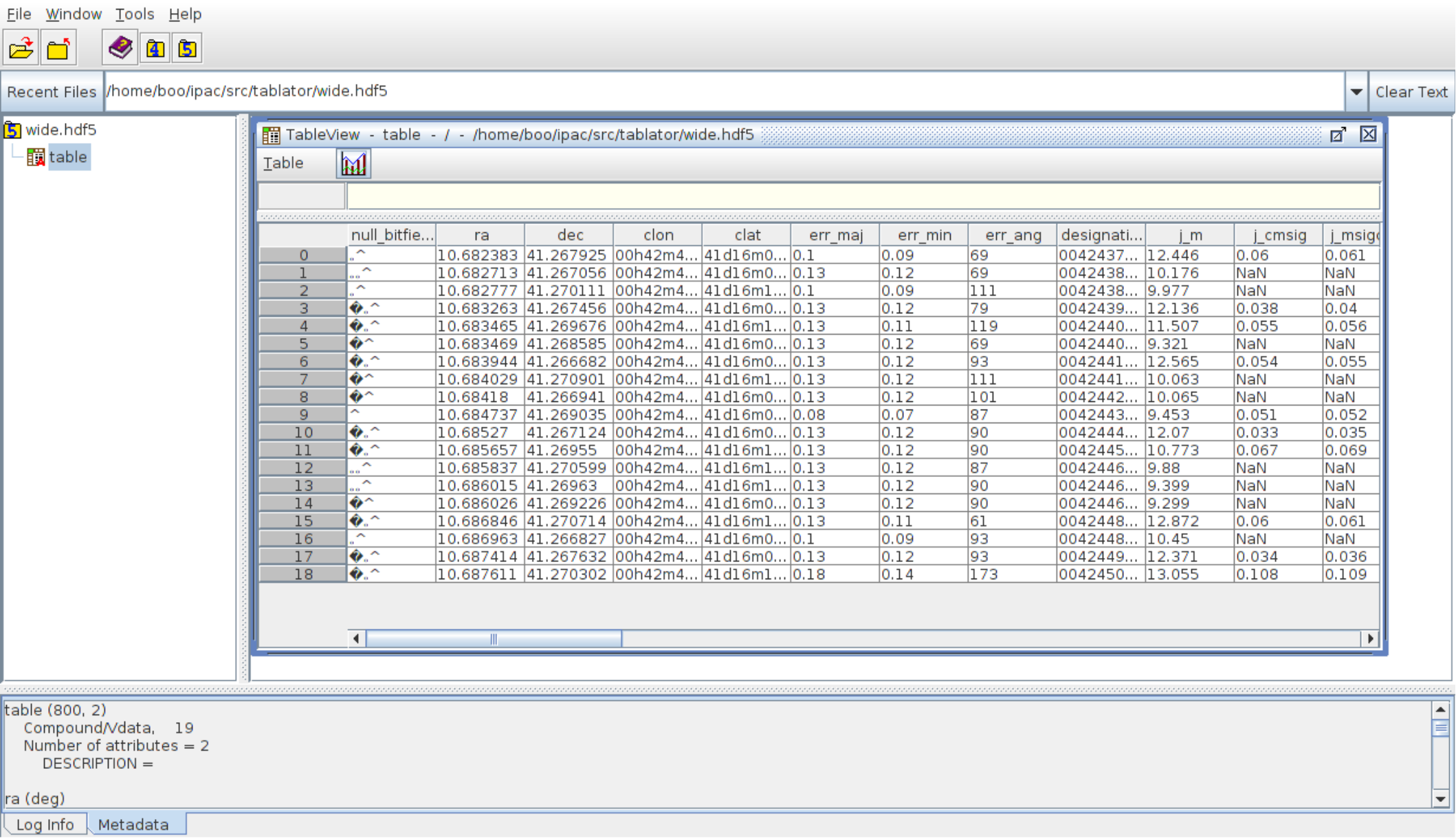
Writing HDF5 Column Data

- The columns map to HDF5 types
  - except Bool
  - no duplicate column names
- Still need to implement the generic column metadata. Only have names for now.

Writing HDF5 Column Data

- The columns map to HDF5 types
  - except Bool
  - no duplicate column names
- Still need to implement the generic column metadata. Only have names for now.
- Writing the data itself is just blitting the internal representation with H5::DataSet::write.
  - extremely fast
    - No endianess flips
    - All columns at once
  - Makes supporting variable length strings tricky

HDFView



	null_bitfie...	ra	dec	clon	clat	err_maj	err_min	err_ang	designati...	j_m	j_cmsig	j_msig
0	„^	10.682383	41.267925	00h42m4...	41d16m0...	0.1	0.09	69	0042437...	12.446	0.06	0.061
1	„„^	10.682713	41.267056	00h42m4...	41d16m0...	0.13	0.12	69	0042438...	10.176	NaN	NaN
2	„^	10.682777	41.270111	00h42m4...	41d16m1...	0.1	0.09	111	0042438...	9.977	NaN	NaN
3	?„^	10.683263	41.267456	00h42m4...	41d16m0...	0.13	0.12	79	0042439...	12.136	0.038	0.04
4	?„^	10.683465	41.269676	00h42m4...	41d16m1...	0.13	0.11	119	0042440...	11.507	0.055	0.056
5	?^	10.683469	41.268585	00h42m4...	41d16m0...	0.13	0.12	69	0042440...	9.321	NaN	NaN
6	?„^	10.683944	41.266682	00h42m4...	41d16m0...	0.13	0.12	93	0042441...	12.565	0.054	0.055
7	?^	10.684029	41.270901	00h42m4...	41d16m1...	0.13	0.12	111	0042441...	10.063	NaN	NaN
8	?^	10.68418	41.266941	00h42m4...	41d16m0...	0.13	0.12	101	0042442...	10.065	NaN	NaN
9	^	10.684737	41.269035	00h42m4...	41d16m0...	0.08	0.07	87	0042443...	9.453	0.051	0.052
10	?„^	10.68527	41.267124	00h42m4...	41d16m0...	0.13	0.12	90	0042444...	12.07	0.033	0.035
11	?„^	10.685657	41.26955	00h42m4...	41d16m1...	0.13	0.12	90	0042445...	10.773	0.067	0.069
12	„„^	10.685837	41.270599	00h42m4...	41d16m1...	0.13	0.12	87	0042446...	9.88	NaN	NaN
13	„„^	10.686015	41.26963	00h42m4...	41d16m1...	0.13	0.12	90	0042446...	9.399	NaN	NaN
14	?^	10.686026	41.269226	00h42m4...	41d16m0...	0.13	0.12	90	0042446...	9.299	NaN	NaN
15	?„^	10.686846	41.270714	00h42m4...	41d16m1...	0.13	0.11	61	0042448...	12.872	0.06	0.061
16	„^	10.686963	41.266827	00h42m4...	41d16m0...	0.1	0.09	93	0042448...	10.45	NaN	NaN
17	?„^	10.687414	41.267632	00h42m4...	41d16m0...	0.13	0.12	93	0042449...	12.371	0.034	0.036
18	?„^	10.687611	41.270302	00h42m4...	41d16m1...	0.18	0.14	173	0042450...	13.055	0.108	0.109