

VizieR & VOUnits

S. Derriere, G. Landais

IVOA Interoperability meeting – Ops2
2022/04/29



□ Units usage in the VO

- Report by Mark Taylor at Nov. 2021 meeting
 - <https://wiki.ivoa.net/internal/IVOA/InterOpNov2021Ops/uc.pdf>
- cds.vizier (Nov. 2021)
 - 422286 columns with units
 - 110883 errors (26.3%)
 - 60 deprecated

| unit | ncol | status | msg |
|---------|-------|------------|---|
| "h:m:s" | 12850 | BAD_SYNTAX | unity parser error at character 1: syntax error |
| "d:m:s" | 12615 | BAD_SYNTAX | unity parser error at character 1: syntax error |
| [-] | 9233 | BAD_SYNTAX | unity parser error at character 1: syntax error |
| [Sun] | 8674 | BAD_SYNTAX | unity parser error at character 1: syntax error |
| % | 3820 | BAD_SYNTAX | unity parser error at character 1: syntax error |
| mW/m2 | 3058 | BAD_SYNTAX | unity parser error at character 5: syntax error |
| [cm/s2] | 2998 | BAD_SYNTAX | unity parser error at character 1: syntax error |

□ Units usage in the VO

- Actions suggested in Nov. 2021 :
 - Review results
 - Improve declared metadata as appropriate
 - Changes to VOUnits standard ?

□ Units in VizieR

- Syntax : Standards for astronomical catalogues 2.0
 - <http://vizier.u-strasbg.fr/doc/catstd-3.2.htx>
 - (used for comparison in VOUnits document)
- ... with some additions and flexibility !
 - Extra symbols supported : **bar**, **Mgeo** (aka **geoMass**), **au**, **μas**...
- « unit » column in METAcot (metadata for columns)
- Used in
 - Web pages
 - Registry VOResource descriptions
 - VOTable output (ok until VOTable 1.3)
- 805128 columns in VizieR (442653 with non-empty units)
- 2208 different expressions for « unit »

| | | | | |
|-------------------------------------|-----------------------|----------|--|---------------------------|
| <input type="checkbox"/> | <input type="radio"/> | recno | | |
| <input checked="" type="checkbox"/> | <input type="radio"/> | RA(ICRS) | | "h:m:s" |
| <input checked="" type="checkbox"/> | <input type="radio"/> | DE(ICRS) | | "d:m:s" |
| <input checked="" type="checkbox"/> | <input type="radio"/> | pmRA | | s/yr |
| <input checked="" type="checkbox"/> | <input type="radio"/> | pmDE | | arcsec/yr |
| <input type="checkbox"/> | <input type="radio"/> | e_RAs | | mas |

□ Compatibility with VOUnits

- VizieR units have extra symbols and use slightly different syntax (grammar)

exponentiation m^{-2} → `m**-2`

multiple / $W/m^2/sr$ → `W.m**-2.sr**-1`

symbols M_{sun} → `solMass`

logarithm $[yr]$ → `log(yr)`

...

- VOTable 1.4 requires VOUnits for contents of `unit=""`
→ need for a new piece of metadata in METAcol :
« `vounit` » column is being added

□ VizieR update unit → vounit

- Documentalists still use « CDS » units for catalogue ReadMe and internal usage
- Need to translate : unit → vounit (PostgreSQL)
 - For legacy metadata (442k columns)
 - In the VizieR ingestion pipeline for new catalogues
- Use of N. Gray's **unity** library for parsing, or M. Taylor's **stilts** for validation : very useful tools!
- **UPDATE METAcOL SET vounit=unit ;** does not work
26.4 % parsing errors on original units : how far can we improve the situation?

□ Improvements

- Syntax of numeric scale factors ($10+35W \rightarrow 1e+35W$), combination&exponents ($W/m2/nm \rightarrow W.m^{*-2}.nm^{*-1}$), logarithm ($[cm^{-2}] \rightarrow \log(cm^{*-2})$), and a few symbols translated
26.4 % errors \rightarrow 11.04 %
- Main issue remaining at this point : « formats »
 - Use of « unit » to indicate the nature of a string, mainly :
 - **6.13 %** are Sexagesimal coordinates : 'h:m:s', 'd:m:s'
 - **0.61 %** are Date, time : 'Y:M:D', JD, MJD, ...

□ Improvements (2)

- Parsing library errors ? Normative grammar too restrictive ? (some symbols in VOUnits Tab.14)

- **2.22 %** **Sun, log(Sun)** should be allowed but

```
$ ./unity -i vounits -gv Sun
0 0/Sun/1
check: all units recognised?      no
```

- **0.93 %** Use of **%**

```
$ ./unity -i vounits -gv %
parse error: units parsing error:
syntax error
```

- **0.65 %** Numerical scale factor in log() log(sym)
e.g. **log(1e-07W)** not allowed by grammar, but
87 different expressions use it in Vizier !

Note : C version of unity fails on extreme numerical scale factors, e.g. 1.0e-46W

□ Improvements (3)

- With these improvements : **26.4 %** errors → **0.5 %**
- With a few more rules to translate :
 - **Rjup, Mjup, Rgeo, Mgeo, ...** (0.20%)
 - **ppm** → **1e-6** (0.06%)
 - **uas** → **uarcsec** (0.04 %)
 - **Lsun** → **solLum** (0.04%)
- And if we ignore warnings of deprecated units (**Angstrom, erg, Gauss, barn**) – 0.06 %

→ **errors below 0.1 % !!** (typos and exotic cases)

□ Conclusions (1)

- A new « vounit » param is being integrated in Vizier, and will be used for
 - Vizier VOTable 1.4 output, for attribute unit="".
 - VOResource description
- We need to address the sexagesimal coordinates
'h:m:s', 'd:m:s'
 - We can have any non-standard symbol for vounit between single quotes " ... but not the : character
 - These are more data formats than units : empty unit+ some xtype ?
 - From

```
<FIELD arraysize="10" datatype="char" name="RAJ2000" ref="J2000" ucd="pos.eq.ra;meta.main" unit="&quot;h:m:s&quot;;">
```
 - To

```
<FIELD arraysize="10" datatype="char" name="RAJ2000" ref="J2000" ucd="pos.eq.ra;meta.main" unit="" xtype="sexagesimal">
```

□ Conclusions (2)

- Erratum for VOUnits standard ? → upgrade validators
 - Support Sun
 - % should be allowed
VOUnit table 2 and 14 not consistent
 - Allow numerical scale factor in log()
function_application: STRING OPEN_P complete_expression CLOSE_P
 - Allow simple numerical factor ? unit='100'
- Make sure we don't break clients relying on units by changing VizieR VOTable output
 - Or have them consume VOTable 1.3
- Work in progress... can we reach 99.9 % compliance ?