The Future of FITS and Other Standardized Astronomical Data Formats

Jessica Mink,[1] Preben Grosbøl,[2] and Rob Seaman[3]

[1]*Smithsonian Astrophysical Observatory, Cambridge, MA, USA;* jmink@cfa.harvard.edu

[2]*European Southern Observatory, Garching b. München, DE*

[3]*University of Arizona Lunar and Planetary Laboratory, Tuscon, AZ, USA*

Abstract. The FITS data standard has served astronomers well for four decades. The original integer image format has been revised to support additional pixel data types, to support world coordinates and other scientific metadata, to include an integrated data compression framework, and to support generalized binary tables, among other features. Over the years, a variety of alternative scientific data standards have been proposed which usually reach only a limited audience specific to a particular project or community. No other format has ever garnered the widespread support of FITS. We will discuss where we can go from here, and how.

1. A Brief FITS History

In its original presentation (Wells & Greisen 1979; Wells et al. 1981), FITS, standing for **F**lexible **I**mage **T**ransport **S**ystem, was primary intended only as a standard data transport format, not as an internal format for data processing. The choice of the Network Byte Order seems reasonable considering the emphasis on the transport of data. Some common processors have used another byte order or floating point format, but as long as the data are translated in or out of the internal processor format only to intake or ouput the data, overhead is small. While FITS has evolved (Mink 2015), the authors still have data from the 1980's and 1990's from Data General, VAX, Sun, early Unix microprocessors, and HP systems which can only be accessed because the internal formats were translated to FITS. The data formats of the Image and Table extensions are also very efficient.

It is reasonable for projects to create their own efficient, internal format data as long as a translation to FITS is available. History shows that projects and image processing systems have a limited lifetime e.g. due to financing or changes in technology.

2. Other Astronomical Data Formats

The metadata and data formats of the FITS Image and Table extensions are very efficient. It is important to note that a change in format can be extremely expensive both in software and administration. Whereas the software for a stable format, like FITS,

may need occasional updates due to compiler changes, a changing data format requires policies for software migration, distribution and version management.

Old readers (without updates) must be able to read a new FITS file without an error, even without being able to understand some new format data. The slogan "Once FITS, always FITS" (Scroggins & Boscoe 2020) works over time in both directions!

If ever one would consider the creation of a new data format for exchange, the following points should be considered:

- Once defined, it must obey the FITS rule: Once FITS always FITS.
- Image and Table format should be retained as they are efficient.
- The World Coordinate System (WCS) should be retained as it is unique.
- It should be easy to navigate through the data stream e.g. to skip over data.

3. Upgrading FITS

Other than data compression (Pence 2002) and radio-specific formats, and marginally documented FITS data and metadata structures used by some processing systems such as IRAF, the community is generally using fairly standard FITS, although with complaints.

The most addressable concerns involve the inflexibility of the FITS header's metadata format. It is clear that the 80-character Header format is inflexible and has significant limitations, but that format cannot be changed significantly without being unreadable to old readers.

One could use the binary table extension for storage of metadata. That and a bit of scaffolding to organize separate extensions and/or tiles in a compressed image would address most of the "modern" data format quibbles. A tile compressed image is about as efficient a storage and transport format as you can get, both in the entropy meaning and for access through an API. Half the battle is encouraging projects to build performant APIs.

Alternatively, it would be possible to create a new FITS extension type which after a standard header could contain a stream of ASCII characters which could contain XML to transport more complex metadata or even more simply, linefeed-terminated FITS-header-structured lines of

```
keyword1 = value1 / comment1<lf>
keyword2 = value2 / comment2<lf>
. . .
keywordn = valuen / commentn<lf>
END<lf>
```

where each element is of an arbitrary length and the list can extend across 2880-byte blocks with a final END line. Multiple extensions could be used to contain different sets of metadata assignments. Either of these would be ignored by FITS readers which were not upgraded to recognize them through their extension headers, which would include

```
XTENSION= 'xxxxxxxxx' / Type of extension
BITPIX  = 8           / Character length in bits
NAXIS   = 1           / Only one axis
NAXIS1  = n           / Length of metadata file in characters
```

FITS readers which don't expect the ASCII could just treat it as an 8-bit binary data file and skip over it.

4. Future Structured Data Standards

The existing Astrophysical Structured Data Format (ASDF (Greenfield et al. 2015)), not be be confused with the Adaptable Seismic Data Format ASDF (Smith et al. 2013) is being used by several large telescope projects. Other groups are looking at AstroHDF (Masters et al. 2012), an implementation of the widely used Hierarchical Data Format 5 or HDF5 (Jenness 2015). Both include multiple kinds of data and metadata in a structured file, and FITS files for exchange can be ingested or produced by either (Price et al. 2015).

So far, as far as sharing metadata standards, such as keyword definitions and updated WCS standards, between projects, there is only discussion.

The IAU Data Representation Working Group was set up several years ago as a node in the structure of the IAU under which the FITS Working Group and other data format standards working groups could reside, but the IAU Data Representation Working Group has never been formalized. Its default chair, the first author of this paper, is nearing retirement and would like to be replaced. The FITS Special Expert Group, which replaced the FITS Working Group, has never been convened in its post-IAUFWG composition. Both groups would benefit from a revision of memberships to ensure that all major projects are represented by active members dealing the data storage and archiving.

The IAU Data Representation Working Group could include a additional Special Expert groups covering VO data formats as IAU Standards, which is already being discussed independently with the IAU, structured data formats such as ASDF and AHDF, and any other formats which it might be useful to standardize in structure and or metadata.

Something will happen to the FITS format and governance over the next years and decades, and changes will be implemented by folks other than the old guard. How can that transition be best mediated? We are thinking about a focused Future of FITS workshop with a basic format of about 50 or so attendees representing specific invited projects. It would be 2 or 2.5 days, with a limited number of presentations and moderated discussions. We would start with a specific plan and exit with a revised plan and outline for a white paper. This would be a good opportunity to bring new folks onto the IAU Working and Expert Group(s) and could be a starting point for other standards which don't work well inside FITS.

Acknowledgments. Some people missed the excitement of past angry discussions of the shortcomings of FITS as those participating in this year's discussion realized that complaints must lead to action on the part of the complainers, but there was great participation from attendees.

References

Greenfield, P., Droettboom, M., & Bray, E. 2015, Astronomy and Computing, 12, 240

Jenness, T. 2015, Astronomy and Computing, 12, 221. 1502.04029

Masters, J., Alexov, A., Folk, M., Hanisch, R., Heber, G., & Wise, M. 2012, in Astronomical Data Analysis Software and Systems XXI, edited by P. Ballester, D. Egret, & N. P. F. Lorente, vol. 461 of Astronomical Society of the Pacific Conference Series, 871

Mink, J. D. 2015, Astronomy and Computing, 12, 128. 1507.03646

Pence, W. D. 2002, in Astronomical Data Analysis II, edited by J.-L. Starck, & F. D. Murtagh, International Society for Optics and Photonics (SPIE), vol. 4847, 444 . URL https://doi.org/10.1117/12.461053

Price, D. C., Barsdell, B. R., & Greenhill, L. J. 2015, Astronomy and Computing, 12, 212. 1505.06421

Scroggins, M., & Boscoe, B. 2020, IEEE Annals of the History of Computing, 42, 42. 1809.09224

Smith, J. A., Bozdag, E., Krischer, L., Lefebvre, M., Lei, W., Podhorszki, N., & Tromp, J. 2013, in AGU Fall Meeting Abstracts, vol. 2013, S33A

Wells, D. C., & Greisen, E. W. 1979, in Image Processing in Astronomy, edited by G. Sedmak, M. Capaccioli, & R. J. Allen, 445

Wells, D. C., Greisen, E. W., & Harten, R. H. 1981, A&AS, 44, 363