Astronomical Data Analysis Software and Systems VI
ASP Conference Series, Vol. 125, 1997
Gareth Hunt and H. E. Payne, eds.

Speculations on the Future of FITS

Donald C. Wells

National Radio Astronomy Observatory,[1] *Charlottesville, VA 22903-2475, E-mail: dwells@nrao.edu*

Abstract. The history and philosophy of FITS are reviewed, with emphasis on the lessons-learned and on the archival requirements. Opinions are offered on the likely outcome of current FITS negotiations, such as the year-2000 problem and the WCS proposal, and on possible subjects of future data interchange format negotiations in astronomy. `BINTABLE` schemas in third-normal-form are advocated. The long-term importance of the `BINTABLE` format as a platform for future layered-convention agreements is stressed.

1. On the Philosophy of FITS (Lessons-learned)

FITS [Flexible Image Transport System] provides a common canonical *language* for talking about astronomical data structures and, as such, it has a profound positive influence on software design practice in astronomy. By negotiating FITS as a *family* of similar data formats, Basic-FITS (1979), random-groups (1980), generalized extensions (1983), `TABLE` (1984), `BINTABLE` (1991) and `IMAGE` (1992), we have minimized the negotiation, documentation and training costs for our community. Our most effective negotiating strategy has been to try to achieve bi-lateral agreements. We include "escape hatches" in our agreements, in places where we expect to negotiate future agreements. The history of FITS shows that it is not possible to fully transmit *meanings.* Instead, the purpose of FITS is agree on the *syntax* of a language for talking about astronomical data, and to agree on the *semantics* in only a limited range of cases. Agreement on syntax permits basic portability and interchange of data, and the users are able to bridge the semantic gaps.

Newcomers to FITS often ask: "Why doesn't FITS have a `VERSION` keyword?" Our answer is: "It does, but the value is always 1.0 by default." The point is that the introduction of a `VERSION` code would be incompatible with the use of FITS as an archival format, because designers of new software would be tempted to support only recent versions. The FITS committees will never knowingly obsolete existing conforming FITS files. This policy is often summarized as "once FITS, always FITS."

Seventeen years of production experience with FITS have demonstrated that only a few actual mistakes were made in the design of Basic-FITS, and that they have not hurt us (yet). A minor mistake is that we specified keyword

[1]The National Radio Astronomy Observatory is a facility of the National Science Foundation, operated under cooperative agreement by Associated Universities, Inc.

EPOCH instead of EQUINOX. A more serious mistake is that DATExxxx='31/12/99' was specified, which exposes the FITS community to the infamous "year-2000" problem; we must correct this within the next three years.[2] The author (one of the original designers) wishes he could change two "mistakes" of style: (1) we should have specified use of SI units more clearly and, in particular, we should have specified radians, the SI auxiliary unit for angles, instead of degrees, and (2) we should have explicitly advocated use of a hierarchical keyword notation, such as the "HISTORY VLACV MAP METHOD='FFT'" notation which appears on line 2/4 of Fig. 1 of Wells, Greisen, & Harten (1981).

2. FITS as an Archival Format

"A data set that is not used by its creator in its archived form is notoriously unreliable."

FITS is not only a way to talk to remote astronomers in the here and now, it is also a way to talk to future astronomers. The FITS standards have been published, and copies will be available in libraries around the world *forever.* The human-readable (self-documenting) headers of FITS, with 60% of the characters *reserved* for comments, complement the published rules of FITS. One alternative interchange format, HDF [Hierarchical Data Format], uses an API [Application Programming Interface] with registered binary tags instead of human-readable self-descriptions in the bitstream. This type of architecture is not as safe as FITS for archival applications because we cannot predict the future in computer languages and operating systems over periods of *decades.* Therefore, *our archival format must always be defined at bitstream level, as FITS is, not by an API.*

3. Repeating Groups Considered Harmful

Our BINTABLE extension is a superb exchange format for normalized databases (sets of related tables). Consider a telescope with multiple detectors operating in parallel, each producing a matrix, each with different dimensionality and WCS parameters. If these detectors are dumped at the same timestamp, should all of the matrices be recorded in the same row of a BINTABLE or should they be recorded as multiple rows with one matrix per row? Only the latter schema is capable of becoming a *normalized* relational database, i.e., of being cast into Third Normal Form, the simplest and most compact schema concept. The first schema (multiple matrices in the same row) is an example of a *repeating group.* Repeating group schemas are harder to design and program, more costly to maintain, and do not support flexible query techniques; the database industry has deprecated them for the past twenty years (Martin 1977, p. 245). Repeating group schemas require that we invent complex conventions to form subscripted column labels for matrix dimensionality, WCS parameters, etc. This complex keyword notation should not be necessary in BINTABLE, because any repeating group schema can be re-designed as a normalized relational schema. Let's apply Occam's razor!

[2] A recent poll of our community has produced the consensus that we will support DATExxxx='1999-12-31', with optional ISO-8601 time notation.

4. FITS Evolution—Work-in-progress

"The purpose of standardization is to aid the creative craftsman, not to enforce the common mediocrity."

Clever pieces of craftsmanship like the `CHECKSUM` proposal (Seaman 1995) can greatly enhance FITS without actually changing it. The author recommends that `CHECKSUM` be implemented in astronomy data systems. The FITS community expects to define and implement a new syntax for `DATExxxx` value strings before 1999-12-31, while agreeing to continue to support the old syntax. We expect to also agree that optional time values can be appended to the date strings. We continue to work toward a celestial coordinates WCS [World Coordinate Systems] agreement. The 25 projections of the sphere onto a 2-D FITS image as specified by Greisen & Calabretta (1996) have been implemented in four different languages (FORTRAN, C/C++, IDL, Java) already. It is likely that we will eventually also agree on spectroscopic and time-series coordinate conventions. Probably we will agree to allow non-printing codes like CR/LF to be used in undefined fields of `TABLE` extensions, in order to make it easier to upload `TABLE` bodies into commercial database software. It appears likely that the `BINTABLE` variable-length vector convention[3] (Cotton et al. 1995) will be widely implemented and used in the future.

5. FITS Evolution—Some Future Possibilities

It is easy to speculate about future FITS agreements—it is much harder to actually negotiate them! The following items are some ideas that the author considers to be possibilities, but which he may or may not support in future negotiations. First, there are a number of ways in which we could agree to "loosen" FITS header syntax, e.g., move the "`=`" around, support lowercase keywords, longer keywords, hierarchical keywords, longer string values, header line continuation convention, etc. We should be *very* cautious about most such header syntax changes, but it is a fact that we could make many of them in such a way as to preserve backward compatibility. We could agree to allow extended character sets (probably the UTF-7/RFC1642 version of ISO-10646/Unicode) in string values of keywords like `OBSERVER` and in `TABLE` extensions. We could agree to support `BITPIX=1`. We could adopt a wide variety of conventions layered on top of `BINTABLE`, such as codings for high performance image compression algorithms, or the Jennings et al. (1995) hierarchical grouping proposal; the author expects that almost all future FITS object types will be layered on `BINTABLE`. We could agree to support `XTENSION='MPEG'` or other MIME-coded types in order to associate such objects with our datasets (a FITS generalized extension is capable of encapsulating *any* other bitstream format). In particular, `XTENSION='JAVA'` might enable us to transmit *portable methods* along with our data objects.

[3] This feature of `BINTABLE` is an especially clever piece of craftsmanship.

6. Has FITS Outlived Its Usefulness?

We need an interchange and archival format more than ever, so the short answer to the question must be “No!” Therefore, the real question is whether we should decide to adopt some other existing format or should negotiate a new format agreement. The author’s opinion is that the potential alternative formats are only slightly stronger than FITS in their areas of strength, and are significantly weaker than FITS in its areas of strength.The costs of re-designing FITS (negotiation, R&D, documentation, retraining, coding support in hundreds of applications) would be enormous. It is very unlikely that the possible gains of re-design could ever be worth all of these costs. Indeed, we would incur most of these costs even if we adopted an existing design from another discipline. Furthermore, it may no longer be possible to negotiate a general interchange and archival format for a community as large, diverse and sophisticated as astronomy now is. Perhaps we were lucky: 1979 was a moment when there were very few vested interests and when several of the largest software projects were still in their startup phases, and were able to adopt FITS as their external canonical form at about the same time. The author expects that these conclusions about the role of FITS will remain true for several more decades.

“We must indeed all hang together, or, most assuredly,
we shall all hang separately.”

References

Cotton, W. D., Tody, D., & Pence, W. D. 1995, A&AS, 113, 159

Greisen, E. W., & Calabretta, M. 1996, Representations of celestial coordinates in FITS[4]

Jennings, D. G., Pence, W. D., & Folk, M. 1995, in ASP Conf. Ser., Vol. 77, Astronomical Data Analysis Software and Systems IV, ed. R. A. Shaw, H. E. Payne, & J. J. E. Hayes (San Francisco: ASP), 229

Martin, J. 1977, Computer Data-Base Organization (Englewood Cliffs, NJ: Prentice-Hall)

Seaman, R. 1995, in ASP Conf. Ser., Vol. 77, Astronomical Data Analysis Software and Systems IV, ed. R. A. Shaw, H. E. Payne, & J. J. E. Hayes (San Francisco: ASP), 247

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

[4] http://fits.cv.nrao.edu/documents/wcs/wcs.all.ps